

# NONCOMMUTATIVE VERSIONS OF THE ARITHMETIC-GEOMETRIC MEAN INEQUALITY

WAFAA ALBAR, MARIUS JUNGE, AND MINGYU ZHAO

Recht and Ré in [21] introduced the noncommutative arithmetic geometric mean inequality (NC-AGM) for matrices with a constant depending on the degree  $d$  and the dimension  $m$ . In this paper we prove AGM inequalities with a dimension-free constant for general operators. We also prove an order version of the AGM inequality under additional hypothesis. Moreover, we show that our AGM inequality almost holds for many examples of random matrices .

## 1. INTRODUCTION

Variations of the arithmetic-geometric mean (AGM) inequality have many applications in analysis and geometry. As pointed out by Ré and Recht in [21], noncommutative versions of the AGM inequalities are relevant to machine learning. In particular, their proof, which employed the classical MacLaurin inequalities, led to improved convergence rate of the of the algorithms in machine learning.

Let us recall the famous MacLaurin inequalities for positive real numbers  $x_1, \dots, x_n$  and the normalized  $d$ -th symmetric sums as

$$S_d = \binom{n}{d}^{-1} \sum_{\substack{\tau \in [n] \\ |\tau|=d}} \prod_{i \in \tau} x_i .$$

where  $1 \leq d \leq n$ . According to the MacLaurin inequalities, we have

$$S_1 \geq \sqrt[2]{S_2} \geq \sqrt[3]{S_3} \geq \dots \geq \sqrt[n]{S_n} .$$

In particular,  $S_1 \geq \sqrt[n]{S_n}$  is the standard AGM inequality. For more details about the classical AGM inequality see [5]. In this paper, we will discuss noncommutative versions of MacLaurin's inequalities. Indeed, we will consider a generalized AGM inequality for the norm and the order. It may come as a surprise to the operator algebra community that these inequalities are motivated by problems in machine learning, stochastic gradient method (see Buttou [2] and the reference there is in [21]), and randomized coordinates descent (see Nesterov [15]). This interesting connection and an overview of known results on this topic can be found in [20] and [21]. In fact, these methods contain an iteration procedure which can be performed with or without replacement samples. Recht and Ré, in [20], study the performance of both. They show that the expected convergence rate without replacement is faster than that with replacement. They proved this result by using a particular AGM inequality.

---

Partially supported by DMS 1501103 and BigData 1447879.

In the effort to generalize the classical AGM inequality to the noncommutative setting, a standard but naive procedure in noncommutative analysis is to replace scalars by operators. Famous examples of this strategy are Cauchy-Schwarz type inequalities for  $C^*$ -modules, Khintchine, and martingale inequalities. (See e.g. LP[12], LPP[13], PXu[18], Narcisse[19], J[6], JXu1[9], JXu2[9]. For a general survey see [18].) Proving these noncommutative extensions often employs a combination of functional analytic and combinatorial methods. In fact, the key results of this paper heavily rely on Pisier's interpretation of Rota's Möbius formulae for partitions.

A NC-AGM inequality would ask whether

$$(1.1) \quad A_1 \cdots A_n \stackrel{?}{\leq} \left( \frac{1}{n} \sum_{j=1}^n A_j \right)^n$$

holds for positive operators  $A_1, \dots, A_n$  on a Hilbert space. (In this context we shall interpret  $x \leq y$  as requiring that  $y - x$  is positive semi-definite.) However, for positive operators  $A$  and  $B$ , the product  $AB$  may not be positive or even self-adjoint, so the inequality (1.1) may not make sense. Inspired by Recht and Ré, we modify (1.1) by replacing the left hand side with the average of all the products of the operators  $A_i$ , which turns out to be self-adjoint. Following the MacLaurin approach, we may now ask whether the AGM inequality holds on average, i.e.

$$(1.2) \quad \frac{1}{n!} \sum_{\sigma \in S_n} A_{\sigma(1)} \cdots A_{\sigma(n)} \stackrel{?}{\leq} \left( \frac{1}{n} \sum_{j=1}^n A_j \right)^n.$$

Unfortunately, we can not prove (1.2) in general. A milder version of (1.2) is to ask for

$$(1.3) \quad \left\| \frac{1}{n!} \sum_{\sigma \in S_n} A_{\sigma(1)} \cdots A_{\sigma(n)} \right\| \stackrel{?}{\leq} \left\| \left( \frac{1}{n} \sum_{j=1}^n A_j \right)^n \right\|,$$

where  $\|x\| = \|x\|_{B(H)}$  refers to the standard operator norm of bounded operators on a Hilbert space  $H$ . The inequality (1.3) is a particular case of the noncommutative MacLaurin inequalities discussed in [20]. Indeed, for fixed  $d$  we may consider the following average product of noncommutative operators of length  $d$ :

$$P_d(A_1, \dots, A_n) = \frac{1}{n \cdots (n-d+1)} \sum_{1 \leq j_1, \dots, j_d \leq n \text{ all different}} A_{j_1} \cdots A_{j_d}.$$

We refer to the example in [21] for the fact that the symmetrization for the operators in the AGM inequality is required. In [20], Ré and Recht posed the following question: Is it true that for positive bounded operators  $A_1, \dots, A_n$  on a Hilbert space one has

$$(1.4) \quad \|P_d(A_1, \dots, A_n)\|^{1/d} \leq \|P_1(A_1, \dots, A_n)\|?$$

They proved that (1.4) holds when  $A_1, \dots, A_n$  are matrices that mutually commute. Moreover, they observed that for operators  $A_1, \dots, A_n$  on an  $m$ -dimensional Hilbert space one has

$$(1.5) \quad \|P_d(A_1, \dots, A_n)\|_{B(\ell_2^m)}^{1/d} \leq m \|P_1(A_1, \dots, A_n)\|.$$

We will prove the AGM inequality for the norm with a constant independent of the dimension  $m$ .

**Theorem 1.1.** *For operators  $A_1, \dots, A_n \geq 0$  on a Hilbert space  $H$ ,*

$$\|P_d(A_1, \dots, A_n)\|^{1/d} \leq d \|P_1(A_1, \dots, A_n)\|.$$

Let us now consider the “order version” of the AGM inequality. Here we add the additional assumption  $\sum A_i = n$ . In order to illustrate the technique we use generally, it is good to start with  $d = 3$ .

**Theorem 1.2.** *Let  $n \geq 6$ . If  $A_1, \dots, A_n$  are self-adjoint operators such that  $\sum_i A_i = n$ . Then  $P_3(A_1, \dots, A_n)^{1/3} \leq 1$ .*

For the proof we consider the mean-zero operators  $a_i := A_i - 1$ . Observe the operators  $a_i$  are self-adjoint and  $\sum_{i=1}^n a_i = 0$ . It follows easily that

$$P_3(A_1, \dots, A_n) = 1 + \binom{3}{1} P_1(a_1, \dots, a_n) + \binom{3}{2} P_2(a_1, \dots, a_n) + \binom{3}{3} P_3(a_1, \dots, a_n).$$

Straightforward computations using  $\sum a_i = 0$  reveal that

$$P_1(a_1, \dots, a_n) = \frac{(n-1)!}{n!} \sum_i a_i = 0$$

$$P_2(a_1, \dots, a_n) = \frac{(n-2)!}{n!} \sum_{i \neq j} a_i a_j = \frac{(n-2)!}{n!} \left( (\sum a_i)^2 - \sum a_i^2 \right) = -\frac{(n-2)!}{n!} \sum a_i^2$$

$$\begin{aligned} P_3(a_1, \dots, a_n) &= \frac{(n-3)!}{n!} \left( \sum_{i \neq j \neq k} a_i a_j a_k \right) \\ &= \frac{(n-3)!}{n!} \left( \left( \sum_{i \neq j \neq k} a_i \right)^3 - \left( \sum_{i=j} a_i^2 \right) \left( \sum_k a_k \right) - \left( \sum_i a_i \right) \left( \sum_{j=k} a_j^2 \right) - \sum_j \sum_{i=k \neq j} a_i a_j a_i + 2 \left( \sum_{i=j=k} a_i^3 \right) \right) \\ &= 2 \frac{(n-3)!}{n!} \left( \sum_i a_i^3 \right). \end{aligned}$$

This leads to the form  $P_3(A_1, \dots, A_n) = 1 - \frac{3}{n(n-1)} \sum a_i^2 + \frac{2}{n(n-1)(n-2)} \sum a_i^3$ . Together with  $\sum a_i^3 \leq \|a_i\| \sum a_i^2 \leq n \sum a_i^2$ , this yields

$$(1.6) \quad P_3(A_1, \dots, A_n) \leq 1 - \frac{3}{n(n-1)} \sum a_i^2 + \frac{2n}{n(n-1)(n-2)} \sum a_i^2.$$

Since  $\frac{2n}{n(n-1)(n-2)} \leq \frac{3}{n(n-1)}$  holds for all  $n \geq 6$ , the right side of (1.6) is at most 1 and we are done. A far-reaching generalization of this idea leads to the following result.

**Theorem 1.3.** *Fix  $n$  and  $d$ . Suppose  $A_1, \dots, A_n$  and  $a_i$  are defined as above,  $\sum_i A_i = n$  and*

$$i) \quad P_1(A_1, \dots, A_n) = \frac{\sum_i A_i}{n} = 1,$$

$$ii) \quad \|(\sum a_j^2)^{\frac{1}{2}}\| \leq \frac{n}{3d}.$$

*Then the AGM inequality holds in the order sense:*

$$P_d(A_1, \dots, A_n) \leq P_1(A_1, \dots, A_n)^d = 1.$$

Note that these techniques work efficiently when  $d$  is very large.

This paper is organized as follows. In section 2, we review the analytic and combinatorial tools needed to prove Theorem 1.2, especially Pisier's interpretation of Rota's results on Möbius transforms for partitions. In section 3, we combine the results from section 2 with Pisier's group construction for partitions in [17] to obtain our key norm and order estimate. In section 4 and 5, a combination of Pisier's partition method and probabilistic results allow "almost AGM" inequalities hold in many different scenarios. We confirm the AGM inequality up to  $\varepsilon$  for many random matrices, in particular Wishart random matrices, more general vector-valued moments of convex bodies, and freely independent operators. We should point out that in contrast to results on averages of random matrices in Ré and Recht in [21], our estimates hold with high probabilities.

## 2. PARTITION AND MÖBIUS FORMULA

We need some definitions from the combinatorial theory of partitions. Let  $\mathbb{P}_d$  be the lattice of all the partitions of  $\{1, \dots, d\}$ . For two partitions  $\sigma$  and  $\pi$ , we write  $\sigma \leq \pi$  if every block of the partition  $\sigma$  is contained in some block of  $\pi$  (i.e., any block of the partition of  $\pi$  can be written as a union of blocks of  $\sigma$ ). In other words,  $\pi$  is a refinement of  $\sigma$ . There are two trivial partitions,  $\dot{0}$  and  $\dot{1}$ , where  $\dot{0}$  is the partition into  $n$  singletons and  $\dot{1}$  is the partition of a single block. For a partition  $\pi$ ,  $\nu(\pi)$  is the number of the blocks of the partition  $\pi$  and  $r_i(\pi)$  is the number of blocks of  $\pi$  with cardinality  $i$  such that  $\sum_{i=1}^d i r_i(\pi) = d$ ; and  $\sum_{i=1}^d r_i(\pi) = \nu(\pi)$ . For more information on partitions, see [1] and [22].

Let us recall some main results on the Möbius function  $\mu$  in [17] which are crucial for our paper.

**Proposition 2.1.** (*Pisier's Möbius inversion formula*) *For any  $d \in \mathbb{N}$  there exists a function  $\mu : \mathbb{P}_d \times \mathbb{P}_d \rightarrow \mathbb{Z}$  such that for every vector space  $V$  and functions  $\phi : \mathbb{P}_d \rightarrow V$  and  $\psi : \mathbb{P}_d \rightarrow V$ , we have the following properties:*

- i) If  $\psi(\sigma) = \sum_{\pi \leq \sigma} \phi(\pi)$ , then  $\phi(\sigma) = \sum_{\pi \leq \sigma} \mu(\pi, \sigma) \psi(\pi)$ ;
- ii) If  $\psi(\sigma) = \sum_{\pi \geq \sigma} \phi(\pi)$ , then  $\phi(\sigma) = \sum_{\pi \geq \sigma} \mu(\sigma, \pi) \psi(\pi)$ ;
- iii) Moreover,  $\forall \sigma \neq \dot{0}$ ,  $\sum_{\dot{0} \leq \pi \leq \sigma} \mu(\pi, \sigma) = 0$ .

The next result provides precise formulas for the Möbius function  $\mu$  in special cases.

**Theorem 2.2.** *The Möbius function satisfies the following properties:*

- i)  $\mu(\dot{0}, \dot{1}) = (-1)^{d-1} (d-1)!$ .
- ii)  $\mu(\dot{0}, \pi) = \prod_{i=1}^d [(-1)^{i-1} (i-1)!]^{r_i(\pi)}$ , and consequently,
- iii)  $\sum_{\pi \in \mathbb{P}_d} |\mu(\dot{0}, \pi)| = d!$ .

If  $\sigma$  is a partition of  $\{1, \dots, d\}$ , then there exists a coordinate function  $f : \{1, \dots, d\} \rightarrow \{1, \dots, \nu(\sigma)\}$  such that  $f^{-1}(t) = A_t$  where each  $A_t$  represents a block in our partition. Note that this coordinate function isn't unique. For every partition  $\sigma$  we can fix an enumeration of the blocks  $f : \{1, 2, \dots, d\} \rightarrow \{1, 2, \dots, |\sigma|\}$  where  $\sigma := \langle j_1, j_2, \dots, j_d \rangle$ . This means  $j_r = j_s$  if and only if  $r, s \in A_{r,s}$  where  $A_{r,s}$  is a block

in  $\sigma = \langle j_1, j_2, \dots, j_d \rangle$ . Using this notation we define the restricted and full partition for elements from an algebra.

**Definition 2.3.** Let  $\mathcal{A}$  be an algebra and  $x_{j_i}^i \in \mathcal{A}$ . The restricted partition is defined by:

$$\langle \sigma \rangle = \sum_{\langle j_1, j_2, \dots, j_d \rangle = \sigma} x_{j_1}^1 \dots x_{j_d}^d.$$

The full partition with elements  $x_{j_i}$  is given by:

$$[\sigma] = \sum_{\pi \geq \sigma} \langle \pi \rangle.$$

The restricted and full partitions, which are denoted as  $\langle \sigma \rangle$  and  $[\sigma]$ , respectively, give expressions for the elements in the given  $B(H)$  according to the algebraic combinatorial partition  $\sigma$ . In order to understand the difference between the definition of restricted partition and full partition, consider the following example.

**Example 2.4.** Let both the numbers of total samples and chosen samples be 3 ( $n = d = 3$ ). Then for the full partition  $[1\ 2\ 3]$ , with the assumption that  $x_j^i = x_j$  we have

$$\begin{aligned} [1\ 2\ 3] &= (\sum x_i^2)(\sum x_i) \\ &= \langle 1\ 2\ 3 \rangle + \langle 1\ 2\ 3 \rangle \\ &= \sum_{i_1=i_2 \neq i_3} x_{i_1}^2 x_{i_3} + \sum_{i_1=i_2=i_3} x_{i_1}^3. \end{aligned}$$

Whereas the restricted partition  $\langle 1\ 2\ 3 \rangle$  is defined as  $\langle 1\ 2\ 3 \rangle = \sum_{i_1=i_2 \neq i_3} x_{i_1}^2 x_{i_3}$ .

We reformulate Pisier's Möbius inversion formula in our context.

**Proposition 2.5.** Let  $x_j^k \in \mathcal{A}$  as above. Then we have

$$(2.1) \quad \langle \pi \rangle = \sum_{\nu \geq \pi} \mu(\pi, \nu) [\nu], \text{ where } [\pi] = \sum_{\nu \geq \pi} \langle \nu \rangle,$$

$$(2.2) \quad \langle \pi \rangle = \sum_{\nu \leq \pi} \mu(\pi, \nu) [\nu], \text{ where } [\pi] = \sum_{\nu \leq \pi} \langle \nu \rangle.$$

Moreover, we have

$$(2.3) \quad \langle \dot{0} \rangle = [\dot{0}] + \sum_{\dot{0} \leq \nu \leq \dot{1}} \mu(\dot{0}, \nu) [\nu].$$

In [17], in order to separate different partition blocks into disjoint subspaces, Pisier uses a trick to embed operators  $x_{ik} \in B(H)$  into  $B(K \otimes H)$  (for another Hilbert space  $K$ ). Our first goal is to modify Pisier's trick by using matrix units.

Consider first the trivial partition that has only one block  $[1, 2, \dots, d]$ . We can write

$$\begin{aligned} \dot{1} = [1, 2, \dots, d] &= \sum x_{i_1}^1 x_{i_2}^2 \dots x_{i_d}^d \\ &= (\sum e_{1i_1} \otimes x_{i_1}^1) \times (\sum e_{i_2 i_2} \otimes x_{i_2}^2) \times \dots \\ &\times (\sum e_{i_{d-1} i_{d-1}} \otimes x_{i_{d-1}}^{d-1}) \times (\sum e_{i_d 1} \otimes x_{i_d}^d). \end{aligned}$$

Now if we have 6 elements and our partition  $\sigma$  has two crossing blocks, one containing  $\{1, 3, 4, 6\}$  and the other containing  $\{2, 5\}$ , then the full partition of  $\sigma$  will be of the form:

$$[\sigma] = \sum_{\substack{i_1=i_3=i_4=i_6 \\ i_2=i_5}} x_{i_1} x_{i_2} x_{i_3} x_{i_4} x_{i_5} x_{i_6}.$$

We rewrite these elements into a tensor form, as follows:

$$\begin{aligned} Z_{i_1} &= e_{1i_1} \otimes 1 \otimes x_{i_1}, & Z_{i_2} &= 1 \otimes e_{1i_2} \otimes x_{i_2} \\ Z_{i_3} &= e_{i_3i_3} \otimes 1 \otimes x_{i_3}, & Z_{i_4} &= e_{i_4i_4} \otimes 1 \otimes x_{i_4} \\ Z_{i_5} &= 1 \otimes e_{i_51} \otimes x_{i_5}, & Z_{i_6} &= e_{i_61} \otimes 1 \otimes x_{i_6}. \end{aligned}$$

With this new notation, we get

$$\begin{aligned} [\sigma] &= \sum_{\substack{i_1=i_3=i_4=i_6 \\ i_2=i_5}} x_{i_1} x_{i_2} x_{i_3} x_{i_4} x_{i_5} x_{i_6} \\ &= \sum_{i_1, i_2, i_3, i_4, i_5, i_6} Z_{i_1} Z_{i_2} Z_{i_3} Z_{i_4} Z_{i_5} Z_{i_6} \\ &= \prod_{j=1}^6 \left( \sum_{i_j} Z_{i_j} \right) = \prod_{j=1}^6 Z_j, \quad (Z_j := \sum_{i_j} Z_{i_j}). \end{aligned}$$

In a more general setting, assume  $\sigma$  has more than one block. Denote  $A_1, \dots, A_{|\sigma|}$  as the blocks of the partition  $\sigma$  with cardinality larger than one.

Then we define

$$(2.4) \quad Z_{j_k}^k \in B(H)^{\otimes |\sigma|} \otimes B(H)$$

as follows:

$$\begin{aligned} \forall k \in A_1, \quad Z_{j_k}^k &= t_{A_1}(j_k) \otimes 1 \otimes \dots \otimes x_{j_k}^k \\ \forall k \in A_2, \quad Z_{j_k}^k &= 1 \otimes t_{A_2}(j_k) \otimes 1 \otimes \dots \otimes x_{j_k}^k \\ \forall k \in A_{|\sigma|}, \quad Z_{j_k}^k &= 1 \otimes \dots \otimes t_{A_{|\sigma|}}(j_k) \otimes x_{j_k}^k, \end{aligned}$$

where

$$t_{A_m}(j_k) = \begin{cases} e_{1j_k} & j_k = \min A_m \\ e_{j_k j_k} & \text{otherwise} \\ e_{j_k 1} & j_k = \max A_m. \end{cases}$$

Here,  $\min A_m$  means the smallest index number and  $\max A_m$  means the largest index number in the partition  $A_m$ . Finally, if  $k$  belongs to singleton block of the partition  $\sigma$ , then we set

$$Z_{j_k}^k = 1 \otimes \dots \otimes 1 \otimes x_{j_k}^k.$$

To sum up, the method places each element into larger spaces, which will allow us to interchange the summation and multiplication as in the above example and the following lemma.

**Lemma 2.6.** *For an arbitrary partition  $\sigma$  for  $d$  elements, we have*

$$[\sigma] = \sum_{i_1, \dots, i_d} Z_{i_1}^1 \dots Z_{i_d}^d.$$

Indeed, this immediately follows from

$$Z_{i_j}^i \cdot Z_{i_k}^k = 0, \text{ if } i_j \neq i_k.$$

Follow Pisier's result in [17]; we deduce the following norm estimate.

**Theorem 2.7.** *For an arbitrary partition  $\sigma$  for  $d$  elements, we have*

$$\|[\sigma]\|_{B(H)} \leq \prod_{k=1}^d \left( \left\| \sum_{j_k} Z_{j_k}^k \right\| \cdot 1_{\sigma_s}(k) + \left\| \sum_{j_k} Z_{j_k}^k \right\| \cdot 1_{\sigma_{ns}}(k) \right).$$

Moreover,  $\|[\sigma]\|_{B(H)} \leq \prod_{k \in \sigma_s} \left\| \sum_{j_k} Z_{j_k} \right\| \times \prod_{k \in \sigma_{ns}} \| (Z_{j_k}) \|$ ,

where  $\| (Z_{j_k}) \| = \max \{ \left\| \sum_{j_{k_1}} Z_{j_{k_1}} Z_{j_{k_1}}^* \right\|^{\frac{1}{2}}, \left\| \sum_{j_{k_p}} Z_{j_{k_p}}^* Z_{j_{k_p}} \right\|^{\frac{1}{2}}, \sup_{j_k} \| Z_{j_k} \| \}$ .

Here  $\sigma_s$  means the set of singletons in the partition  $\sigma$ , and  $\sigma_{ns}$  means the set of non-singleton elements in the partition  $\sigma$ . The functions  $1_{\sigma_{ns}}(k), 1_{\sigma_s}(k)$  represent the characteristic functions, i.e.

$$1_{\sigma_{ns}}(k) = \begin{cases} 1 & k \in \sigma_{ns} \\ 0 & \text{otherwise} \end{cases}, \quad 1_{\sigma_s}(k) = \begin{cases} 1 & k \in \sigma_s \\ 0 & \text{otherwise} \end{cases}$$

*Proof.* Taking the norm for the full partition, we have

$$\begin{aligned} \|\sigma\| &= \left\| \sum_{\pi \geq \sigma} \langle \pi \rangle \right\| = \left\| \sum_{\langle j_1, \dots, j_d \rangle \geq \sigma} x_{j_1}^1 \cdots x_{j_d}^d \right\| \\ (2.5) \quad &= \left\| \sum_{j_1, j_2, \dots, j_d} Z_{j_1}^1 \cdots Z_{j_d}^d \right\| \end{aligned}$$

$$\begin{aligned} (2.6) \quad &= \left\| \prod_{k \in \sigma_s} \sum_{j_k} Z_{j_k}^k \cdot \prod_{k \in \sigma_{ns}} \sum_{j_k} Z_{j_k}^k \right\| \\ &\leq \left\| \prod_{k \in \sigma_s} \sum_{j_k} Z_{j_k}^k \right\| \cdot \left\| \prod_{k \in \sigma_{ns}} \sum_{j_k} Z_{j_k}^k \right\| \\ &\leq \prod_{k \in \sigma_s} \left\| \sum_{j_k} Z_{j_k}^k \right\| \cdot \prod_{k \in \sigma_{ns}} \left\| \sum_{j_k} Z_{j_k}^k \right\|. \end{aligned}$$

The equality (2.5) comes from Lemma 2.6. The equality (2.6) follows from the definition of  $Z_{j_k}^k$ , which means it allows us to perform summation first and then multiplication. Next,

$$\begin{aligned} \|[\sigma]\|_{B(H)} &\leq \prod_{k \in \sigma_s} \left\| \sum_{j_k} Z_{j_k} \right\| \cdot \prod_{k \in \sigma_{ns}} \left\| \sum_{j_k} Z_{j_k} \right\| \\ &\leq \prod_{k \in \sigma_s} \left\| \sum_{j_k} Z_{j_k} \right\| \times \prod_{k \in A_m \subset \sigma_{ns}} \left\| \sum_{j_k} Z_{j_k} \right\| \cdot (1_{\min A_m} + 1_{\max A_m} + 1_{\text{mid } A_m}) \\ &\leq \prod_{k \in \sigma_s} \left\| \sum_{j_k} Z_{j_k} \right\| \times \\ &\quad \prod_{k \in A_m \subset \sigma_{ns}} \left( \left\| \sum_{j_k} Z_{j_k} \right\| \cdot 1_{\min A_m} + \left\| \sum_{j_k} Z_{j_k} \right\| \cdot 1_{\max A_m} + \left\| \sum_{j_k} Z_{j_k} \right\| \cdot 1_{\text{mid } A_m} \right) \end{aligned}$$

$$\begin{aligned}
&\leq \prod_{k \in \sigma_s} \left\| \sum_{j_k} Z_{j_k} \right\| \times \\
&\quad \prod_{k \in A_m \subset \sigma_{ns}} \left( \left\| \sum_{j_k} Z_{j_k} Z_{j_k}^* \right\|^{\frac{1}{2}} \cdot 1_{\min A_m} + \left\| \sum_{j_k} Z_{j_k}^* Z_{j_k} \right\|^{\frac{1}{2}} \cdot 1_{\max A_m} + \sup_{j_k} \|Z_{j_k}\| \cdot 1_{\text{mid } A_m} \right) \\
&\leq \prod_{k \in \sigma_s} \left\| \sum_{j_k} Z_{j_k} \right\| \times \prod_{k \in \sigma_{ns}} \| (Z_{j_k}) \|,
\end{aligned}$$

where  $\| (Z_{j_k}) \| = \max \{ \left\| \sum Z_{j_{k_1}} Z_{j_{k_1}}^* \right\|^{\frac{1}{2}}, \left\| \sum Z_{j_{k_p}}^* Z_{j_{k_p}} \right\|^{\frac{1}{2}}, \sup_{j_k} \|Z_{j_k}\| \}$ .  $\blacksquare$

The next corollary states the norm estimate in  $B(H)$  rather than in  $B(K \otimes H)$ . For simplicity we replace  $x_{i_k}^k$  by  $x_{i_k}$ .

**Corollary 2.8.** *If  $\sigma$  is a partition and  $x_{j_k}$  is a self-adjoint operator for arbitrary  $k \in \{1, \dots, d\}$ , then*

$$\|[\sigma]\|_{B(H)} \leq \prod_{k \in \sigma_s} \left\| \sum x_{j_k} \right\| \cdot \prod_{k \in \sigma_{ns}} \left\| \sum x_{j_k}^2 \right\|^{\frac{1}{2}}$$

*Proof.* We need to discuss two cases:

(i) For  $k \in \sigma_s$ ,  $\left\| \sum_j Z_{j_k} \right\| = \left\| \sum_j 1 \otimes \dots \otimes x_{j_k} \right\| = \|1 \otimes \dots \otimes \sum x_{j_k}\| = \left\| \sum x_{j_k} \right\|$ .

(ii) For  $A_m \in \sigma_{ns}$ ,

$$\begin{aligned}
\left\| \sum Z_{j_{k_1}} Z_{j_{k_1}}^* \right\|^{\frac{1}{2}} &= \left\| \sum [1 \otimes \dots \otimes e_{1j_{k_1}} \otimes \dots \otimes x_{j_{k_1}}] \cdot [1 \otimes \dots \otimes e_{j_{k_1}1} \otimes \dots \otimes x_{j_{k_1}}^*] \right\|^{\frac{1}{2}} \\
&= \left\| \sum 1 \otimes \dots \otimes e_{11} \otimes \dots \otimes x_{j_{k_1}} x_{j_{k_1}}^* \right\|^{\frac{1}{2}} \\
(2.7) \quad &= \|1 \otimes \dots \otimes \sum x_{j_{k_1}} x_{j_{k_1}}^*\|^{\frac{1}{2}} = \left\| \sum x_{j_{k_1}} x_{j_{k_1}}^* \right\|^{\frac{1}{2}} = \left\| \sum x_{j_{k_1}}^2 \right\|^{\frac{1}{2}}.
\end{aligned}$$

and

$$\begin{aligned}
\left\| \sum Z_{j_{k_p}}^* Z_{j_{k_p}} \right\|^{\frac{1}{2}} &= \left\| \sum [1 \otimes \dots \otimes e_{1j_{k_p}} \otimes \dots \otimes x_{j_{k_p}}^*] \cdot [1 \otimes \dots \otimes e_{j_{k_p}1} \otimes \dots \otimes x_{j_{k_p}}] \right\|^{\frac{1}{2}} \\
&= \left\| \sum 1 \otimes \dots \otimes e_{11} \otimes \dots \otimes x_{j_{k_p}}^* x_{j_{k_p}} \right\|^{\frac{1}{2}} = \|1 \otimes \dots \otimes \sum x_{j_{k_p}}^* x_{j_{k_p}}\|^{\frac{1}{2}} \\
(2.8) \quad &= \left\| \sum x_{j_{k_p}}^* x_{j_{k_p}} \right\|^{\frac{1}{2}} = \left\| \sum x_{j_{k_p}}^2 \right\|^{\frac{1}{2}}.
\end{aligned}$$

For the middle term, we have

$$\begin{aligned}
\sup_{k \in \{k_2, \dots, k_{p-1}\}} \sup_{j_k} \|Z_{j_k}\| &= \sup_{k \in \{k_2, \dots, k_{p-1}\}} \sup_{j_k} \|Z_{j_k}^* Z_{j_k}\|^{\frac{1}{2}} = \sup_{k \in \{k_2, \dots, k_{p-1}\}} \sup_{j_k} \|x_{j_k}^* x_{j_k}\|^{\frac{1}{2}} \\
(2.9) \quad &\leq \sup_{k \in A_m} \left\| \sum x_{j_k}^2 \right\|^{\frac{1}{2}}.
\end{aligned}$$

Combining (i) and (ii) finishes the proof.  $\blacksquare$

### 3. AGM INEQUALITY FOR THE NORM AND FOR THE ORDER

In this section we prove the AGM inequality for the norm and for the order. We need the following lemma which handles positive or self-adjoint operators  $\{x_{i_k}\}$  in a  $C^*$ -algebra  $\mathcal{A}$ .

**Lemma 3.1.** (i) *If  $x_{j_k} \geq 0$ , then  $\left\| \sum x_{j_k}^2 \right\|^{\frac{1}{2}} \leq \left\| \sum x_{j_k} \right\|$ .*

(ii) *If  $x_{j_k}$  are self-adjoint, then  $\left\| \sum x_{j_k}^2 \right\|^{\frac{1}{2}} = \left\| \left( \sum x_{j_k}^2 \right)^{\frac{1}{2}} \right\|$ .*



*Proof.* (i) Indeed, we have

$$\begin{aligned} \left\| \sum x_{j_k}^2 \right\|^{\frac{1}{2}} &= \left\| \sum x_{j_k}^{\frac{1}{2}} x_{j_k} x_{j_k}^{\frac{1}{2}} \right\|^{\frac{1}{2}} \\ &\leq \left( \left\| \sum x_{j_k} \right\|^{\frac{1}{2}} \cdot \left\| \sum x_{j_k} \right\| \cdot \left\| \sum x_{j_k} \right\|^{\frac{1}{2}} \right)^{\frac{1}{2}} \\ &= \left\| \sum x_{j_k} \right\|. \end{aligned}$$

(ii) Holds trivially using  $\|x^2\| = \|x\|^2$ , for  $x = (\sum x_{j_k}^2)^{\frac{1}{2}}$ . ■

**3.1. AGM inequality for the norm.** Now we have done all the preparation to prove the NC-AGM inequality for the norm.

**Theorem 3.2.** *Suppose  $x_1, \dots, x_n$  are positive operators in  $B(H)$ . Then*

$$(3.1) \quad \|P_d(x_1, \dots, x_n)\|_{B(H)}^{1/d} \leq d \|P_1(x_1, \dots, x_n)\|_{B(H)}.$$

*Proof.* From Corollary 2.8 and Lemma 3.1, we deduce that for a given arbitrary partition  $\sigma$  and positive elements  $x_{j_k} = x_j$ , we have

$$\|[\sigma]\|_{B(H)} \leq \left\| \sum x_j \right\|^d.$$

Recall identity 2.3 from Proposition 2.5:

$$(3.2) \quad \langle 1, \dots, d \rangle = [1, \dots, d] + \sum_{v \not\geq \dot{0}} \mu(\dot{0}, \nu)[\nu], \quad \text{where } \sum_{v \not\geq \dot{0}} |\mu(\dot{0}, \nu)| = d! - 1.$$

Taking the norm of both sides of the equality (3.2) we get

$$\begin{aligned} \|\langle 1, \dots, d \rangle\|_{B(H)} &= \|[1, \dots, d] + \sum_{v \not\geq \dot{0}} \mu(\dot{0}, \nu)[\nu]\|_{B(H)} \\ &\leq \|[1, \dots, d]\|_{B(H)} + \sum_{v \not\geq \dot{0}} |\mu(\dot{0}, \nu)| \|\nu\|_{B(H)} \\ &\leq \left\| \sum x_j \right\|_{B(H)}^d + (d! - 1) \left\| \sum x_j \right\|_{B(H)}^d \\ &\leq d! \left\| \sum x_j \right\|_{B(H)}^d \\ &= d! n^d \left\| \frac{1}{n} \sum x_j \right\|_{B(H)}^d \\ &= d! n^d \|P_1(x_1, \dots, x_n)\|_{B(H)}. \end{aligned}$$

Thus,

$$\|P_d(x_1, \dots, x_n)\|_{B(H)} \leq \frac{d! n^d (n-d)!}{n!} \|P_1(x_1, \dots, x_n)\|_{B(H)}.$$

Denote  $C(n, d) := \frac{d! n^d (n-d)!}{n!}$ , and for fixed  $d$  define  $f(n) := \sum_{i=0}^{d-1} \log \frac{n}{n-i}$ . Then

$$\begin{aligned} C(n, d) &= \frac{d! n^d (n-d)!}{n!} = \frac{d! n^d}{n(n-1)(n-2) \cdots (n-d+1)} \\ &= d! \cdot \frac{n}{n} \cdot \frac{n}{n-1} \cdot \frac{n}{n-2} \cdots \frac{n}{n-d+1} \\ &= d! \cdot \exp(f(n)). \end{aligned}$$

Since  $f(n)$  is a decreasing function in  $n$ ,  $C(n, d)$  is also a decreasing function with respect to the variable  $n$ . From the definition of  $d$ , we know  $n \geq d$ , so  $\max_{n \geq d} C(n, d) = C(d, d) = d^d$ .  $\blacksquare$

**3.2. AGM inequality for the order.** Recall that the average product is defined by:

$$P_d(x_1, x_2, \dots, x_n) = \frac{(n-d)!}{n!} \sum_{\langle \sigma \rangle = \dot{0}} x_{i_1} \dots x_{i_d}.$$

**Lemma 3.3.** *Let  $\{x_i\}$  be a finite family of positive operators in  $B(H)$  which satisfy the condition  $\sum_{i=1}^n x_i = n$ . If  $a_i := x_i - 1$  then*

$$(3.3) \quad P_d(x_1, x_2, \dots, x_n) = 1 + \sum_{k=1}^d \binom{d}{k} P_k(a_1, a_2, \dots, a_n).$$

*Proof.* This lemma can be proved by two methods. The first method is by induction which is left to the reader. For the convenience of the reader we give the second proof, using the binomial identity. Then we have

$$\begin{aligned} P_d(x_1, \dots, x_n) &= \frac{(n-d)!}{n!} \sum_{\langle \sigma \rangle = \dot{0}} x_{i_1} \dots x_{i_d} \\ &= \frac{(n-d)!}{n!} \sum_{\langle \sigma \rangle = \dot{0}} (a_{i_1} + 1)(a_{i_2} + 1) \dots (a_{i_d} + 1) = 1 + \sum_{k=1}^d \lambda_k P_k(a_1, \dots, a_n). \end{aligned}$$

Let  $x_1 = x_2 = \dots = x_n = t$ , where  $t = a + 1$ . Then

$$P_d(x_1, \dots, x_n) = t^d = (1+a)^d = 1 + \sum_{k=1}^d \binom{d}{k} a^k,$$

which implies that  $\lambda_k = \binom{d}{k}$ , so  $P_d(x_1, \dots, x_n) = 1 + \sum_{k=1}^d \binom{d}{k} P_k(a_1, \dots, a_n)$ .  $\blacksquare$

In Theorem 1.2 for  $d=3$ , we deduce that each term in  $P_3(x_1, \dots, x_n)$  has an upper bound of some scalar multiple of  $\sum a_i^2$ . For  $d > 3$ , we need the following lemma.

**Lemma 3.4.** *If  $\{x_i\}, \{a_i\}$  are defined as above, then*

$$\max_i \|a_i\| \leq \left\| \sum a_i^2 \right\|^{\frac{1}{2}} \leq \left\| \sum x_i^2 \right\|^{\frac{1}{2}}.$$

*In particular,  $\|a_i\|^k \leq n^k \left\| \frac{1}{n^2} \sum_i x_i^2 \right\|^{\frac{k}{2}}$ .*

*Proof.* Since we have  $a_j^2 \leq \sum a_i^2$ ,

$$\|a_i\| = \|a_i^2\|^{\frac{1}{2}} \leq \left\| \sum a_i^2 \right\|^{\frac{1}{2}}.$$

Moreover, for each  $a_i$ , we have  $x_i = a_i + 1$ . Thus

$$\sum x_i^2 = \sum a_i^2 + n \geq \sum a_i^2$$

This finishes the proof.  $\blacksquare$

Note that for a partition with  $d = 3$ , the proof of the AGM inequality in the order sense was easily done in the introduction. However, the proof is much more complicated for  $d \geq 4$ . The complication comes from crossing partitions, so we need the following useful known lemma [16].

**Lemma 3.5.** *Assume  $a, b \in B(H)$  and  $t \geq 0$ . Then*

$$\begin{aligned} (1) \quad & -(a^*a + b^*b) \leq a^*b + b^*a \leq a^*a + b^*b \\ (2) \quad & ab + b^*a^* \leq t^2aa^* + t^{-2}b^*b \end{aligned}$$

To prove (1), we start by observing  $(a+b)^*(a+b), (a-b)^*(a-b) \geq 0$ . This directly gives  $-(a^*a + b^*b) \leq a^*b + b^*a$  and  $a^*b + b^*a \leq a^*a + b^*b$ . It is clear that (2) is a special case of (1), using the assumptions that  $a = ta^*$  and  $b = t^{-1}b^*$  for the upper bound of (1).

The two previous lemmas will help in establishing our result for general case of the AGM inequality for the order. For convenience, we will write  $A_i := \sum_i Z_i$  where  $Z_i$  is defined as at the beginning of Section 3.1. We now provide upper and lower bounds for  $P_d(a_{i_1}, \dots, a_{i_n})$ .

**Lemma 3.6.** *If  $\{a_i\}$  and  $\{x_i\}$  are defined as above, then for  $S = \|\sum x_i^2\|^{1/2}$*

$$-\frac{(n-d)!}{n!}d! S^{d-2} \sum a_i^2 \leq P_d(a_1, a_2, \dots, a_n) \leq \frac{(n-d)!}{n!}d! S^{d-2} \sum a_i^2.$$

*Proof.* From Proposition 2.5, we know

$$\frac{n!}{(n-d)!}P_d(a_1, a_2, \dots, a_n) = \langle \dot{0} \rangle_d = [\dot{0}]_d + \sum_{\dot{0} \lesssim \nu \leq 1} \mu(\dot{0}, \nu)[\nu]_d.$$

We will prove first the case when  $\mu(\dot{0}, \nu) \geq 0$ . We will obtain an upper bound for the sum  $[\nu]_d$  by introducing  $[\bar{\nu}]_d$  as the following:

$$\begin{aligned} [\nu]_d &= \sum_{\langle i_1, i_2, \dots, i_d \rangle \geq \nu} a_{i_1} a_{i_2} \cdots a_{i_d}, \\ [\bar{\nu}]_d &:= \sum_{\langle i_1, i_2, \dots, i_d \rangle \geq \nu} a_{i_d} a_{i_{d-1}} \cdots a_{i_1}. \end{aligned}$$

Here the  $\bar{\nu}$  can be viewed as the transposition of the partition  $\nu$ . By Theorem 2.2, we have  $\mu(\dot{0}, \pi) = \prod_{i=1}^d [(-1)^{i-1}(i-1)!]^{r_i(\pi)}$ . So  $\mu(\dot{0}, \nu) = \mu(\dot{0}, \bar{\nu})$ . Thus, we can sum these two items together.

**Claim:** For every partition  $\nu$  and  $S = \|\sum x_i^2\|^{1/2}$  we have

$$(3.4) \quad -2 S^{d-2} \sum a_i^2 \leq [\nu]_d + [\bar{\nu}]_d \leq 2 S^{d-2} \sum a_i^2.$$

The idea here is to use our modification of Pisier's trick for these two partitions. Recall that  $Z_{i_1} = e_{1i_1} \otimes a_{i_1}$  is for the first component in the partition,  $Z_{i_j} = e_{jj} \otimes a_{i_j}$  is for the elements in the middle of the partition, and  $Z_{i_d} = e_{i_d 1} \otimes a_{i_d}$  is for the

last element in the partition. Then we have

$$\begin{aligned}
[\nu]_d + [\bar{\nu}]_d &= \sum_{\langle i_1, i_2, \dots, i_d \rangle \geq \nu} a_{i_1} a_{i_2} \cdots a_{i_d} + a_{i_d} a_{i_{d-1}} \cdots a_{i_1} \\
&= \sum_{i_1} Z_{i_1} \cdots \sum_{i_d} Z_{i_d} + \sum_{i_d} Z_{i_d}^* \cdots \sum_{i_1} Z_{i_1}^* \\
&= A_1 \cdots A_d + A_d^* \cdots A_1^*.
\end{aligned}$$

By applying Lemma 3.5 with  $S = \|\sum x_i^2\|^{1/2}$ ,

$$\begin{aligned}
(3.5) \quad [\nu]_d + [\bar{\nu}]_d &= A_1 \cdots A_d + A_d^* \cdots A_1^* \\
&\leq t^2 A_1 A_1^* + t^{-2} A_d^* \cdots A_2^* A_2 \cdots A_d \\
&\leq t^2 A_1 A_1^* + t^{-2} \prod_{j=2}^{d-1} \|A_j^* A_j\| A_d^* A_d \\
&\leq t^2 A_1 A_1^* + t^{-2} \prod_{j=2}^{d-1} \|A_j\|^2 A_d^* A_d \\
(3.6) \quad &\leq t^2 A_1 A_1^* + t^{-2} \prod_{j=2}^{d-1} \|\sum a_j^2\| A_d^* A_d
\end{aligned}$$

$$\begin{aligned}
(3.7) \quad &\leq t^2 A_1 A_1^* + t^{-2} \|\sum a_i^2\|^{d-2} A_d^* A_d \\
&\leq \|\sum a_i^2\|^{d/2-1} (A_1 A_1^* + A_d^* A_d) \leq 2 \sum a_i^2 S^{d-2}.
\end{aligned}$$

Indeed, if our partition contains the singleton then  $[\nu]_d + [\bar{\nu}]_d$  is already zero. Hence we may assume there are no singletons in our partition as it also can be noticed in inequality (3.7). Indeed, if the index is a singleton in partition  $\nu$ , then it is controlled by the summation norm  $\|\sum a_i\|$  which is zero by our construction. On the other hand, if the index is in a non-singleton block, then by Theorem 2.8, it is controlled by the square norm  $\|\sum a_i^2\|$ . Therefore, in both cases,  $\|A_i\|$  is controlled by the square norm of  $a_i$ . To get inequality (3.6), we may apply the norm equality as in equality (2.7) from section 2. For the inequality (3.7), we use Lemma (3.5) by choosing  $t^2 = S^{d/2-1}$ . Then we have

$$\begin{aligned}
\frac{n!}{(n-d)!} P_d(a_1, a_2, \dots, a_n) &= [\dot{0}]_d + \sum_{\dot{0} \not\leq \nu \leq \dot{1}} \mu(\dot{0}, \nu) [\nu]_d = \sum_{\dot{0} \not\leq \nu \leq \dot{1}} \mu(\dot{0}, \nu) [\nu]_d \\
&= \sum_{\mu(\dot{0}, \nu) \geq 0} \mu(\dot{0}, \nu) [\nu]_d + \sum_{\mu(\dot{0}, \nu) \leq 0} \mu(\dot{0}, \nu) [\nu]_d \\
&= \frac{1}{2} \left( \sum_{\mu(\dot{0}, \nu) \geq 0} \mu(\dot{0}, \nu) [\nu]_d + \sum_{\mu(\dot{0}, \bar{\nu}) \geq 0} \mu(\dot{0}, \bar{\nu}) [\bar{\nu}]_d \right) \\
&\quad + \frac{1}{2} \left( \sum_{\mu(\dot{0}, \nu) \leq 0} \mu(\dot{0}, \nu) [\nu]_d + \sum_{\mu(\dot{0}, \bar{\nu}) \leq 0} \mu(\dot{0}, \bar{\nu}) [\bar{\nu}]_d \right) \\
\frac{n!}{(n-d)!} P_d(a_1, a_2, \dots, a_n) &\leq \sum_{\mu(\dot{0}, \nu) \geq 0} \mu(\dot{0}, \nu) S^{d-2} \sum a_i^2 - \sum_{\mu(\dot{0}, \nu) \leq 0} \mu(\dot{0}, \nu) S^{d-2} \sum a_i^2 \\
&= \sum |\mu(\dot{0}, \nu)| S^{d-2} (\sum a_i^2) = d! S^{d-2} \sum a_i^2.
\end{aligned}$$

For the lower bound, the proof is similar to the one above replacing  $A_1$  by  $-A_1$ .  $\blacksquare$

**Theorem 3.7.** (*AGM inequality for the order*) Fix  $n$  and  $d$ . Let  $x_1, \dots, x_n$  be self-adjoint operators such that  $\sum_i x_i = n$  and  $a_i = x_i - 1$  as above. Assume the following conditions hold:

- i)  $P_1(x_1, \dots, x_n) = \frac{\sum_i x_i}{n} = 1$ ,
- ii)  $\|(\sum x_i^2)^{\frac{1}{2}}\| \leq \frac{n}{3d}$ .

Then the AGM inequality holds in the order sense:

$$P_d(x_1, x_2, \dots, x_n) \leq \left( \frac{\sum_i x_i}{n} \right)^d = 1.$$

*Proof.* According to Lemma 3.4, we have  $\|\sum a_i^2\|^{1/2} \leq \frac{n}{3d}$ . Using this upper bound for the average of noncommutative operators  $a_i$  with the identity (3.3) where  $S = \|\sum x_i^2\|^{1/2} \leq \Delta n$  and let  $\Delta := \frac{1}{3d}$ , we have

$$\begin{aligned} P_d(x_1, x_2, \dots, x_n) &= 1 + \sum_{k=1}^d \binom{d}{k} P_k(a_1, a_2, \dots, a_n) \\ &= 1 - \binom{d}{2} \frac{(n-2)!}{n!} (\sum a_i^2) + \sum_{k=3}^d \binom{d}{k} P_k(a_1, a_2, \dots, a_n) \\ &\leq 1 - \binom{d}{2} \frac{(n-2)!}{n!} (\sum a_i^2) + \sum_{k=3}^d \binom{d}{k} \frac{(n-k)!}{n!} k! \Delta^{k-2} n^{k-2} (\sum a_i^2). \end{aligned}$$

Now we need the following condition:

$$(3.8) \quad \binom{d}{2} \frac{(n-2)!}{n!} \stackrel{?}{\geq} \sum_{k=3}^d \binom{d}{k} \frac{(n-k)!}{n!} k! \Delta^{k-2} n^{k-2}.$$

Simplifying the right hand side gives

$$\begin{aligned} \sum_{k=3}^d \binom{d}{k} \frac{(n-k)!}{n!} k! \Delta^{k-2} n^{k-2} &= \sum_{k=3}^d \frac{d!}{(d-k)! k!} \frac{(n-k)! k!}{n!} \Delta^{k-2} n^{k-2} \\ &= \sum_{k=3}^d \frac{d!}{(d-k)!} \frac{(n-k)!}{n!} \Delta^{k-2} n^{k-2} \\ (3.9) \quad &= \frac{1}{n(n-1)} \sum_{k=3}^d \frac{d!}{(d-k)!} \frac{n^{k-2}}{(n-2) \cdots (n-k+1)} \Delta^{k-2}. \end{aligned}$$

Fix  $k$ , and denote  $f(n) := \frac{n^{k-2}}{(n-2) \cdots (n-k+1)}$ . Then, by taking the logarithm, we have  $g(n) := \log f(n) = \sum_{i=2}^{k-1} \log \frac{n}{n-i}$ . Observe that  $g(n)$  is a decreasing function and thus  $f(n)$  is a decreasing function as well. Therefore, we get the inequality:

$$(3.10) \quad \frac{n^{k-2}}{(n-2) \cdots (n-k+1)} \leq \frac{d^{k-2}}{(d-2) \cdots (d-k+1)}.$$

We continue the calculation in (3.9) with the help of inequality (3.10), we have

$$\begin{aligned}
\sum_{k=3}^d \binom{d}{k} \frac{(n-k)!}{n!} k! \Delta^{k-2} n^{k-2} &= \frac{1}{n(n-1)} \sum_{k=3}^d \frac{d!}{(d-k)!} \frac{n^{k-2}}{(n-2) \cdots (n-k+1)} \Delta^{k-2} \\
&\leq \frac{1}{n(n-1)} \sum_{k=3}^d \frac{d!}{(d-k)!} \frac{d^{k-2}}{(d-2) \cdots (d-k+1)} \Delta^{k-2} \\
&\leq \frac{1}{n(n-1)} \sum_{k=3}^d d(d-1) d^{k-2} \Delta^{k-2} \\
&= \frac{d(d-1)}{n(n-1)} \frac{d\Delta(1-(d\Delta)^{d-2})}{1-d\Delta} \leq \frac{d(d-1)}{n(n-1)} \frac{d\Delta}{1-d\Delta}.
\end{aligned}$$

With our choice of  $\Delta = \frac{1}{3d}$  we deduce indeed  $\frac{d(d-1)}{n(n-1)} \frac{d\Delta}{1-d\Delta} \leq \binom{d}{2} \frac{(n-2)!}{n!}$  and this completes the proof.  $\blacksquare$

#### 4. AGM INEQUALITY FOR RANDOM MATRICES

In this section, we prove a version of the NC-AGM inequality for random matrices. We start with a deviation inequality. Let us use the norm  $\|X\|_p = (E\|X\|_{B(H)}^p)^{1/p}$  defined for a random variable  $X : \Omega \rightarrow B(H)$ .

**Proposition 4.1.** *Let  $\{a_i\}$  be a family of self-adjoint random operators. Let  $\varepsilon > 0$ ,  $p \geq 2$ ,  $p_d = \frac{p}{d}$  and  $x_i = a_i + 1$ . Define*

$$(i) \quad \varepsilon_p := \left\| \left\| \frac{1}{n} \sum a_i - E \frac{1}{n} \sum a_i \right\| \right\|_p,$$

$$(ii) \quad \delta_p := \frac{1}{n} \left\| \left\| (\sum a_i^2)^{1/2} \right\| \right\|_p,$$

$$(iii) \quad \gamma_p := \max(\varepsilon_p, \delta_p).$$

Assume  $\sum_i E a_i = 0$  and  $\gamma_p \leq \frac{1}{3d}$  and  $\varepsilon = 3d\gamma_p$

Then,  $\|P_d(x_1, \dots, x_n) - EP_d(x_1, \dots, x_n)\|_{p_d} \leq \varepsilon$ .

*Proof.* From the assumption above, we get that

$$\left\| \left\| \left( \frac{1}{n} \sum a_i \right) \right\| \right\|_p = \varepsilon_p.$$

Fix a partition  $\nu$ . According to Theorem (2.7) and by using Hölder's inequality we have that

$$\begin{aligned}
E\|\nu\|_\infty^{p_d} &\leq E \left( \left\| (\sum a_i^2)^{1/2} \right\|_\infty^{(d-|\nu_s|)p_d} \left\| (\sum a_i) \right\|_\infty^{|\nu_s|p_d} \right) \\
&= E \left( \left\| (\sum a_i^2)^{1/2} \right\|_\infty^{\frac{(d-|\nu_s|)p_d d}{d}} \cdot \left\| (\sum a_i) \right\|_\infty^{\frac{|\nu_s|p_d d}{d}} \right)
\end{aligned}$$

$$\begin{aligned}
&\leq \left( E(\|(\sum a_i^2)^{1/2}\|_\infty^{p_d}) \right)^{\frac{p_d(d-|\nu_s|)}{p_d^d}} \left( E\|(\sum a_i)\|_\infty^{p_d} \right)^{\frac{p_d|\nu_s|}{p_d^d}} \\
&= \left( E(\|(\sum a_i^2)^{1/2}\|_\infty^p) \right)^{\frac{p_d(d-|\nu_s|)}{p}} \left( E\|(\sum a_i)\|_\infty^p \right)^{\frac{p_d|\nu_s|}{p}} \\
&= \left( \left\| \sum a_i^2 \right\|_p^{1/2} \right)^{p_d(d-|\nu_s|)} \left( \left\| \sum a_i \right\|_p \right)^{p_d|\nu_s|} \\
&= (\delta_p \cdot n)^{p_d(d-|\nu_s|)} (\varepsilon_p \cdot n)^{p_d|\nu_s|} \\
&= \delta_p^{p_d(d-|\nu_s|)} \varepsilon_p^{p_d|\nu_s|} n^p = \delta_p^{p_d(d-|\nu_s|)} \varepsilon_p^{p_d|\nu_s|} n^{p_d}.
\end{aligned}$$

Since  $\gamma_p = \max(\delta_p, \varepsilon_p)$ ,

$$(4.1) \quad \left\| [\nu] \right\|_{p_d} = (E\|[\nu]\|_\infty^{p_d})^{\frac{1}{p_d}} \leq \gamma_{p_d}^d \cdot n^d$$

By using our definition of  $\gamma_p$  and the upper bound for inequality (4.1) we obtain

$$\begin{aligned}
&(E\|P_k(a_1, \dots, a_n) - EP_k(a_1, \dots, a_n)\|_\infty^{p_d})^{1/p_d} \\
&\leq \frac{(n-k)!}{n!} \sum |\mu(0, \nu)| \left( E(\|[\nu] - E[\nu]\|_\infty^{p_d}) \right)^{1/p_d} \\
&\leq \frac{(n-k)!}{n!} \sum |\mu(0, \nu)| \cdot 2(E\|[\nu]\|_\infty^{p_d})^{1/p_d} \\
(4.2) \quad &\leq 2 \frac{(n-k)!}{n!} k! \gamma_{p_d k}^k n^k.
\end{aligned}$$

From the above we will have

$$\begin{aligned}
&\left\| P_d(x_1, \dots, x_n) - EP_d(x_1, \dots, x_n) \right\|_{p_d} \\
&= \left\| \sum_{k=1}^d \binom{d}{k} (P_k(a_1, \dots, a_n) - EP_k(a_1, \dots, a_n)) \right\|_{p_d} \\
&= (E\| \sum_{k=1}^d \binom{d}{k} (P_k(a_1, \dots, a_n) - EP_k(a_1, \dots, a_n)) \|_\infty^{p_d})^{1/p_d} \\
&\leq \sum_{k=1}^d \binom{d}{k} (E\|P_k(a_1, \dots, a_n) - EP_k(a_1, \dots, a_n)\|_\infty^{p_d})^{1/p_d} \\
&\leq 2 \sum_{k=1}^d \binom{d}{k} \frac{(n-k)!}{n!} k! \gamma_{p_d k}^k \cdot n^k = 2 \sum_{k=1}^d \frac{d!}{k!(d-k)!} \frac{(n-k)!}{n!} k! \gamma_{p_d k}^k \cdot n^k \\
(4.3) \quad &\leq 2 \sum_{k=1}^d \frac{d!(n-k)!n^k}{(d-k)!n!} \gamma_p^k.
\end{aligned}$$

Recall the definition  $\gamma_{p_d k} = \max(\delta_{p_d k}, \varepsilon_{p_d k})$ . Each  $\delta_{p_d k}, \varepsilon_{p_d k}$  is increasing since  $L_{p_d k}$  is defined as probability space which is norm increasing in probability measure. Thus  $\gamma_{p_d k} \leq \gamma_{p_d} = \gamma_p, \forall 1 \leq k \leq d$ , which justifies the last inequality (4.3). Let  $f(n) = \frac{d!(n-k)!n^k}{(d-k)!n!}$ . This function is a decreasing function in  $n$ , so  $f(d) = \max f(n) =$

$d^k$ . Then we have

$$\begin{aligned} & \|P_d(x_1, \dots, x_n) - EP_d(x_1, \dots, x_n)\|_{p_d} \\ & \leq 2 \sum_{k=1}^d (d \cdot \gamma_p)^k = 2 \cdot d \cdot \gamma_p \frac{(1 - (d \cdot \gamma_p)^d)}{1 - d \cdot \gamma_p} \leq 2 \cdot \frac{d \cdot \gamma_p}{1 - d \cdot \gamma_p} \leq \varepsilon. \end{aligned}$$

The last inequality follows from  $d \cdot \gamma_p \leq \frac{\varepsilon}{1-\varepsilon/2}$ . ■

We now present conditions for positive random operators  $\{x_i\}$  where  $a_i = x_i - 1$ . Note that for  $A := \sum_{i=1}^n \frac{a_i}{n}$ , we have

$$E\|A - EA\|_p = E\left\|\left(\frac{\sum x_j}{n}\right) - E\left(\frac{\sum x_j}{n}\right)\right\|_p.$$

Therefore, whenever we control the  $x_i$ 's, we control the  $a_i$ 's.

**Lemma 4.2.** *Let  $\{x_i\}$  be a family of self-adjoint random operators. Then*

$$\left\|\left(\sum (x_i - 1)^2\right)^{1/2}\right\|_p \leq 6 \left\|\left(\sum x_i^2\right)^{1/2}\right\|_p$$

*Proof.* Observe that  $\left\|\left(\sum x_i^2\right)^{1/2}\right\|_p = \left\|\sum x_i \otimes e_{i,1}\right\|_p$  is given by the column norm. Define operators  $\phi : C_n(B(H)) \rightarrow C_n(B(H))$  and  $\Phi : C_n \rightarrow C_n$  such that  $\Phi(\alpha_i) = (\frac{1}{n} \sum_j \alpha_i)_j$  where  $\phi = \Phi \otimes Id$ . Then it is easy to check that  $\|\Phi\|_{cb} = \|\phi\|_{cb} \leq 1$ .

Indeed

$$\left\|\sum_{j=1}^n e_{j,1} \otimes \phi(y_i)_j\right\| = \left\|\sum_{j=1}^n e_{j,1} \otimes \left(\frac{1}{n} \sum_{i=1}^n y_i\right)\right\| = \frac{1}{\sqrt{n}} \left\|\sum_{j=1}^n y_j\right\| \leq \left\|\sum_i e_{1,i} \otimes y_i\right\|.$$

Denote  $z_i := x_i - Ex_i$ , so  $\left\|\sum_i e_{1,i} \otimes (Id + \Phi)(z_i)\right\| \leq 2 \left\|\sum_i e_{1,i} \otimes x_i\right\|$ . Also,

$$\begin{aligned} (Id + \phi)(z_i) &= x_i - Ex_i + \frac{1}{n} \sum x_i - \frac{1}{n} \sum Ex_i = x_i - 1 - Ex_i + \frac{1}{n} \sum x_i, \\ (x_i - 1) &= (Id + \phi)(z_i) + Ex_i - \frac{1}{n} \sum x_i. \end{aligned}$$

By triangle inequality, we can get

$$\begin{aligned} & \left\|\sum (x_i - 1) \otimes e_{i,1}\right\| \\ & \leq \left\|\sum (Id + \phi)(z_i) \otimes e_{i,1}\right\| + \left\|\sum Ex_i \otimes e_{i,1}\right\| + \left\|\sum_j \left(\frac{1}{n} \sum x_i\right) \otimes e_{j,1}\right\| \\ & \leq 2 \left\|\sum z_i \otimes e_{i,1}\right\| + \left\|\sum Ex_i \otimes e_{i,1}\right\| + \left\|\sum_j \left(\frac{1}{n} \sum x_i\right) \otimes e_{j,1}\right\| \\ & \leq 2 \left\|\sum (x_i - Ex_i) \otimes e_{i,1}\right\| + \left\|\sum Ex_i \otimes e_{i,1}\right\| + \left\|\sum_j \left(\frac{1}{n} \sum x_i\right) \otimes e_{j,1}\right\| \\ & \leq 2 \left\|\sum x_i \otimes e_{i,1}\right\| + 3 \left\|\sum Ex_i \otimes e_{i,1}\right\| + \left\|\frac{1}{n} \sum x_i\right\| \\ & \leq 2 \left\|\left(\sum x_i^2\right)^{\frac{1}{2}}\right\| + 3 \left\|\sum x_i \otimes e_{i,1}\right\| + \left\|\left(\sum x_i^2\right)^{\frac{1}{2}}\right\| \leq 6 \left\|\left(\sum x_i^2\right)^{\frac{1}{2}}\right\|. \end{aligned}$$



The second-to-last inequality  $\|\sum E x_i \otimes e_{i,1}\| \leq \|\sum x_i \otimes e_{i,1}\|$  follows from the fact that conditional expectation from  $E : L_\infty(\Omega, B(H)) \rightarrow B(H)$  is a complete contraction. The inequality

$$\left\| \frac{1}{n} \sum x_i \right\| \leq \left\| \left( \sum x_i^2 \right)^{\frac{1}{2}} \right\|$$

is true by the Cauchy-Schwarz inequality.  $\blacksquare$

Thanks to Theorem 4.1 and Lemma 4.2 we obtain the following deviation result.

**Theorem 4.3.** *Let  $p \geq 2$ ,  $p_d := \frac{p}{d}$ , and  $\{x_i\}$  be a random family of positive operators such that  $E x_i = 1$ . Define*

$$(i) \quad \varepsilon_p := \left\| \frac{1}{n} \sum x_i - E \frac{1}{n} \sum x_i \right\|_p,$$

$$(ii) \quad \delta_p := \frac{1}{n} \left\| \left( \sum x_i^2 \right)^{1/2} \right\|_p,$$

$$(iii) \quad \gamma_p := \max(\varepsilon_p, 4\delta_p).$$

If  $3d \cdot \gamma_p \leq 1$  then

$$\|P_d(x_1, \dots, x_n) - E P_d(x_1, \dots, x_n)\|_{p_d} \leq 3d \cdot \gamma_p.$$

**Corollary 4.4.** *If in addition  $\{x_i\}$  are matrix-valued i.i.d. Then*

$$\|P_d(x_1, \dots, x_n)\|_{p_d} \leq 1 + 3d \cdot \gamma_p.$$

*Proof.* Since  $x_i$ 's are matrix-valued i.i.d, then  $E(P_d(x_1, \dots, x_n)) = P_d(E x_1, \dots, E x_n)$ . Moreover, for  $\varepsilon := 3d \cdot \gamma_p$  by (ii) in the above Theorem 4.3, we have

$$\|(\sum E(x_i)^2)^{1/2}\| = \left\| \sum E(x_i) \otimes e_{i,1} \right\|_{C_n \otimes B(H)} \leq \left\| \sum x_i \otimes e_{i,1} \right\|_{C_n \otimes B(H)} \leq \delta_p \cdot n.$$

Then we can use Theorem 3.7 for  $E(x_i)$ 's and the classical AGM inequality (here  $\delta_p \leq \frac{1}{4}\gamma_p \leq \frac{1}{4d} < \frac{1}{3d}$ ).

$$\begin{aligned} E(P_d(x_1, \dots, x_n)) &\leq P_1(E x_1, \dots, E x_n) \\ &= \sum \frac{E x_i}{n} = 1. \end{aligned}$$

Using the upper bound above and Theorem 4.3, we have the required inequality.

$$\|P_d(x_1, \dots, x_n)\|_{p_d} \leq \|P_1(E x_1, \dots, E x_n)\|_{p_d} + \varepsilon \leq 1 + \varepsilon. \quad \blacksquare$$

#### 4.1. Application to Log concave measures.

In this section we want to study random AGM inequalities for log-concave measures.

**Definition 4.5.** *A Borel measure  $\mu$  on  $n$ -dimensional Euclidean space  $\mathbb{R}^n$  is called logarithmically concave (or log-concave) if for any compact subsets  $A$  and  $B$  of  $\mathbb{R}^n$  and  $0 \leq \lambda \leq 1$  we have*

$$\mu(\lambda A + (1 - \lambda)B) \geq \mu(A)^\lambda \mu(B)^{(1-\lambda)}.$$

Let us recall the isotropic measure  $\mu$  in  $\mathbb{R}^n$ .

**Definition 4.6.** *The isotropic measure  $\mu$  is the measure which satisfies*

$$\int_{\mathbb{R}^n} |\langle \theta, x \rangle|^2 d\mu(x) = L_\mu \|\theta\|^2,$$

for all  $\theta \in \mathbb{R}^n$  where  $L_\mu$  is denoted as isotropic constant.

Also let us recall Rosenthal's inequality, which will be used frequently in this section.

**Theorem 4.7** (Rosenthal inequality [10]). *Let  $A_i$  be a fully independent sub-algebra over  $N$  where  $N \subset M$  and  $M$  is a von Neumann algebra, and  $1 \leq p < \infty$ . Let  $x_i \in L_p(A_i)$  with  $E_N(x_i) = 0$ . Then*

$$\left\| \sum_{i=1}^n x_i \right\|_p \leq C \max \left\{ \sqrt{p} \left\| \sum_{i=1}^n E_N(x_i^* x_i + x_i x_i^*)^{1/2} \right\|_p, p \left( \sum_{i=1}^n \|x_i\|_p^p \right)^{1/p} \right\}.$$

**Theorem 4.8.** *Let  $n, d \in \mathbb{N}$ ,  $p \geq 2$ . Let  $(\mathbb{R}^d, \mu)$  be log-concave Borel measure  $\mu$  in isotropic position on  $\mathbb{R}^d$  with constant  $L$ . Define random variable  $y : \mathbb{R}^d \rightarrow \mathbb{R}^d$  by  $y(\omega) = \frac{\omega}{\sqrt{L}}$  where  $\omega \in \mathbb{R}^d$ . Let  $y_i$  be independent copies of  $y$ . Then  $x_i(\omega) := |y_i(\omega)\rangle\langle y_i(\omega)|$  is a  $d \times d$  random matrix satisfying*

- (i)  $\forall i, E x_i = 1$ ,
- (ii)  $\left\| \sum (x_i - E x_i) \right\|_p \leq \gamma_p \cdot n$ ,
- (iii)  $\left\| \sum x_i^2 \right\|_p^{1/2} \leq \gamma_p \cdot n$ ,

$$\text{where } \gamma_p = \begin{cases} p^{1/2} \sqrt{\frac{d}{n}} + p^{5/2} \frac{d}{n} & p \geq \ln n \text{ or} \\ 2C \sqrt{\ln d} \delta^{1/2} & d \leq \frac{n}{\ln n^5} \\ 2C (\ln n)^3 \delta & d \geq \frac{n}{\ln n^5} \end{cases}$$

- (iv) *Moreover, assume  $\gamma_p \leq (1 - \frac{2}{2+\varepsilon}) \frac{1}{d}$ ,  $\varepsilon \geq 0$ , and  $p_k := \frac{p}{k}$ . Then the following hold.*

- $\left\| P_k(x_1, \dots, x_n) - E P_k(x_1, \dots, x_n) \right\|_{p_k} \leq \varepsilon$ .
- *The AGM inequality holds*

$$\left\| P_k(x_1, \dots, x_n) \right\|_{p_k} \leq (1 + 2\varepsilon).$$

*Proof.* We apply Rosenthal's inequality for  $q \geq p$  to  $x_i - 1$  instead of  $x_i$ . Let us introduce the norm in the space  $L_q(S_q)$  where  $S_q$  is the Schatten class,

$$|x|_q := (E \|x_i\|_{S_q}^q)^{\frac{1}{q}} = \left( \int \|x(\omega)\|_q^q d\mu \right)^{\frac{1}{q}}.$$

So, we have

$$\begin{aligned} & \left\| \sum (x_i - E x_i) \right\|_q \\ & \leq c \max \left\{ \sqrt{q} \left\| \sum E((x_i - 1)^*(x_i - 1) + (x_i - 1)(x_i - 1)^*) \right\|_{\frac{q}{2}}^{\frac{1}{2}}, q \left( \sum |x_i - E x_i|_q^q \right)^{\frac{1}{q}} \right\} \\ & \leq c \sqrt{q} \left\| \sum E((x_i - 1)^*(x_i - 1) + (x_i - 1)(x_i - 1)^*) \right\|_{\frac{q}{2}}^{\frac{1}{2}} + c q \left( \sum |x_i - E x_i|_q^q \right)^{\frac{1}{q}} \\ & \leq 2c \sqrt{q} \left\| \sum E(x_i - 1)^2 \right\|_{\frac{q}{2}}^{\frac{1}{2}} + c q \left( \sum |x_i - E x_i|_q^q \right)^{\frac{1}{q}} \\ & \leq 2c \sqrt{q} \left\| \sum E x_i^2 \right\|_{\frac{q}{2}}^{\frac{1}{2}} + 2c q n^{1/q} |x_1|_q \end{aligned}$$

By Rosenthal's inequality, we need to separately estimate the two terms of the right side. We denote

$$(4.4) \quad \text{I} = |(\sum Ex_i^2)^{1/2}|_q \text{ and } \text{II} = |x_1|_q.$$

We claim that (ii) holds for  $\gamma_q$  and  $Ex_i^2 \leq dEx_i \leq cd \cdot 1$ . Using Borel inequality (see [14] where  $\|\cdot\|$  is seminorm), we have

$$(E\|y\|_X^q)^{\frac{1}{q}} \leq C_q E\|y\|_X \leq C_q (E\|y\|_X^2)^{\frac{1}{2}}.$$

Recall that  $E\|y\|^2 = \sum_{i=1}^d E|\langle e_i, \frac{\omega}{\sqrt{L}} \rangle|^2 = d$ . So, we have for  $x_i := x_1 = |y\rangle\langle y|$

$$\begin{aligned} \langle \theta, Ex_1^2 \theta \rangle &= E\langle \theta, y \rangle \langle y, y \rangle \langle \theta, y \rangle = E\|y\|^2 |\langle \theta, y \rangle|^2 \\ &\leq (E\|y\|^4)^{\frac{1}{2}} (E|\langle \theta, y \rangle|^4)^{\frac{1}{2}} \\ &\leq C_4^4 E\|y\|^2 E(|\langle \theta, y \rangle|^2) = C_4^4 \cdot d \|\theta\|^2. \end{aligned}$$

i.e.  $Ex_i^2 \leq dEx_i \leq cd \cdot 1$ . This implies  $\|(\sum Ex_i^2)^{1/2}\|_q \leq C \cdot d^{1/2+1/q}$  which proves our claim for (I). For (II), note that the  $q$ -norm is defined to be  $|x|_q = (E\text{tr}|x|^q)^{\frac{1}{q}}$ . Let's first take  $q = m$  be an integer. We have

$$\begin{aligned} x_i^m &= |y_i\rangle\langle y_i|^m = y_i \langle y_i, y_i \rangle \cdots \langle y_i, y_i \rangle y_i \\ &= y_i \|y_i\|^{2(m-1)} \langle y_i. \end{aligned}$$

Then, by using the Borel inequality (see [14]), we have

$$\begin{aligned} E\text{tr}(x_i^m) &= E\text{tr}(y_i \|y_i\|^{2(m-1)} \langle y_i) = E(\|y_i\|_2^{2m}) \\ &\leq (C \cdot 2m)^{2m} (E\|y\|_2^2)^{1/2})^{2m} \\ &\leq (C \cdot 2m)^{2m} d^m. \end{aligned}$$

So we get the inequality  $|x|_m \leq (C \cdot 2m)^2 d$  for arbitrary integer  $m$ . Then for any real number  $q$ , we can find an integer  $m$ , such that  $m \leq q \leq m+1$ , and by interpolation between  $m$  and  $m+1$ , we get

$$(4.5) \quad |x|_q \leq (C \cdot 2q)^2 d.$$

Thanks to (4.5), we can now prove condition (iii).

$$\begin{aligned} \left\| \left( \sum x_i^2 \right)^{1/2} \right\|_q &\leq \left\| \sum x_i^2 \right\|_{\frac{q}{2}}^{1/2} \leq \left( \sum \left\| x_i^2 \right\|_{\frac{q}{2}} \right)^{1/2} \\ &\leq \left( \sum \|x_i\|_q \right)^{1/2} = \sqrt{n} \|x_1\|_q \\ (4.6) \quad &\leq \sqrt{n} |x_1|_q \leq \sqrt{n} d (C \cdot 2q)^2. \end{aligned}$$

Combining (I) and (II) we obtain

$$\begin{aligned} \left\| \sum (x_i - 1) \right\|_q &\leq \tilde{c}(qnd)^{1/2} d^{1/q} + cqCn^{1/q} q^2 d \\ &= \tilde{C}(qn)^{1/2} d^{\frac{1}{2} + \frac{1}{q}} + C' n^{1/q} q^3 d. \end{aligned}$$

And then divide each term by  $n$ , we have

$$\begin{aligned}
\frac{\|\sum(x_i - Ex_i)\|_q}{n} &\leq C(q, d, n) := \left(\frac{q}{n}\right)^{1/2} d^{\frac{1}{2} + \frac{1}{q}} + n^{\frac{1}{q}-1} q^3 d \\
&= \left(\frac{q}{n}\right)^{1/2} d^{\frac{1}{2} + \frac{1}{q}} + \frac{qd}{n} q^2 n^{\frac{1}{q}} \\
&= d^{1/q} \left( q^{1/2} \left(\frac{d}{n}\right)^{1/2} + q^3 \frac{d}{n} \frac{n^{1/q}}{d^{1/q}} \right) \\
&= d^{1/q} \left( (q)^{1/2} \left(\frac{d}{n}\right)^{1/2} + q^3 \left(\frac{d}{n}\right)^{1-1/q} \right).
\end{aligned}$$

If we denote  $\frac{d}{n} = \delta$ , then

$$\begin{aligned}
\frac{\|\sum(x_i - Ex_i)\|_q}{n} &\leq d^{\frac{1}{q}} (q^{\frac{1}{2}} \delta^{\frac{1}{2}} + q^3 \delta^{1-\frac{1}{q}}) \\
&= d^{1/q} q^{1/2} \delta^{1/2} (1 + q^{5/2} \delta^{1/2-1/q}).
\end{aligned}$$

Now our goal is to find  $\hat{\gamma}_q = \inf_{q \geq q_0} d^{1/q} q^{1/2} \delta^{1/2} (1 + q^{5/2} \delta^{1/2-1/q})$  by optimization over  $q$  where  $q_0 \geq 2$ . Define  $f(q, \delta) := q^{5/2} \delta^{1/2-1/q}$  and consider

$$g := \ln f(q, \delta) = \frac{5}{2} \ln q + \left(\frac{1}{2} - \frac{1}{q}\right) \ln \delta,$$

with derivative  $g' = \frac{5}{2} \frac{1}{q} + \frac{1}{q^2} \ln \delta$ . The critical point for  $f(q, \delta)$  is  $q(\delta) = \frac{2}{5} \ln \frac{1}{\delta}$ . Since  $f(q, \delta)$  is a convex function then it has no more than one minimum point which is  $q(\delta)$ . Then we have to consider the following cases for the choices of  $q$ ,

- (1)  $q_0 \leq \ln d \leq q_1$  where  $q_1 = (\frac{1}{\delta})^{1/5}$ .
- (2)  $\ln d \leq q_0 \leq \ln n$
- (3)  $\ln d \leq \ln n \leq q_0$

This can be done by using optimization over  $q$  for the term  $d^{1/q} q^{1/2} \delta^{1/2}$ . For the first case, we choose  $q = \ln d$  and  $C(q, \delta) = 2C\sqrt{\ln d} \delta^{1/2}$  where  $f(q, \delta) \leq 1$ . We also calculate  $q_1$  which represents the upper bound for our choice of  $q$  from  $q^{5/2} \delta^{1/2} = 1$ . For the second case, if  $(\frac{n}{d})^{1/5} \geq \ln \frac{n}{d}$ , then we simply choose  $q = \ln n$ . This leads to  $\frac{d}{n} \asymp \frac{1}{\ln n^4} \leq \frac{1}{\ln n^5}$ . We can summarize the cases in the following

$$\hat{\gamma}_q = \begin{cases} q^{1/2} \sqrt{\frac{d}{n}} + q^{5/2} \frac{d}{n} & q \geq \ln n \quad \text{or} \\ 2C\sqrt{\ln d} \delta^{1/2} & d \leq \frac{n}{\ln n^5} \\ 2C(\ln n)^3 \delta & d \geq \frac{n}{\ln n^5} \end{cases}$$

We apply the estimate for  $q \geq p$  and appeal to Theorem 4.3 and Corollary 4.4 to deduce the AGM inequality. ■

**4.2. Wishart random variable matrices.** Let us recall the definition of Wishart random matrices. Let  $[g_{r,s}^i]$  is a family of  $d \times m$  Gaussian random matrices such that  $i \in [1, n]$ ,  $r \in [1, d]$  and  $s \in [1, m]$ . Define  $G_i = \frac{1}{\sqrt{m}} [g_{r,s}^i]$  and  $x_i = G_i G_i^*$ . We call the matrices  $x_i$   $d \times d$  Wishart random matrices. Then we have  $Ex_i = EG_i G_i^* = I$  which implies that  $\sum_{i=1}^n Ex_i = nI$ . In this section we assume that  $m \geq n$ . Let us list some useful lemmas which will be used in the main theorem. Each of these lemmas proves one of the conditions of Theorem 4.3 separately.

**Lemma 4.9.** *Let  $\varepsilon_{q,m,n,d} = \left(\frac{\sqrt{d}+\sqrt{m}}{\sqrt{m}}\right)^2 \frac{q}{\sqrt{n}}$ . Those  $d \times d$  Wishart random matrices  $\{x_i\}$  from above satisfy*

$$(4.7) \quad \frac{1}{n} \left\| \left( \sum x_i^2 \right)^{1/2} \right\|_q \leq \varepsilon_{q,m,n,d}.$$

*Proof.* Denote  $A = \frac{1}{\sqrt{m}} \sum_{r,s} g_{r,s} e_{r,s}$ . Then for all  $h \in H$ , and  $x = AA^*$

$$\begin{aligned} E(h, x^2 h) &= E(h, |AA^*|^2 h) = E(h, AA^* AA^* h) = E(AA^* h, AA^* h) \\ &= E\|AA^* h\|^2 \leq E(\|A\|_{op}^2 \cdot \|A^* h\|^2) \leq E\|A\|_{op}^2 \cdot E\|A^* h\|^2. \end{aligned}$$

Note that  $E\|A^* h\|^2 = E(h, A^* A h) = \|h\|^2$ . Using Chevet's inequality [4],

$$(4.8) \quad E\|A\| = E\left\| \sum_{r=1}^d \sum_{s=1}^m g_{r,s} e_r \otimes e_s \right\|_{X \otimes Y} \leq E\left\| \sum_{s=1}^m g_{r,s} e_s \right\| + E\left\| \sum_{r=1}^d g_{r,s} e_r \right\|,$$

where  $X = l_2^m$  and  $Y = l_2^d$ . We deduce that if  $A = \frac{1}{\sqrt{m}} \sum g_{r,s}^i e_r \otimes e_s$  then by using Kahane's inequality we have that

$$(4.9) \quad (E\|A\|_{op}^2)^{1/2} \leq \sqrt{2} \left( \frac{\sqrt{d} + \sqrt{m}}{\sqrt{m}} \right) =: C(d, m)$$

Therefore

$$\|x_i\|_2 = (E\|x_i\|_{op}^2)^{1/2} \leq C(d, m).$$

For  $q \geq 2$

$$\begin{aligned} \left\| \left( \sum x_i^2 \right)^{1/2} \right\|_q &= \left\| \sum x_i^2 \right\|_{q/2}^{1/2} \leq \left( \sum \|x_i^2\|_{q/2} \right)^{1/2} \\ &\leq \left( \sum \|x_i\|_q^2 \right)^{1/2} \leq \sqrt{n} \|x_i\|_q = \sqrt{n} [(E\|A\|^{2q})^{1/2q}]^2 \\ (4.10) \quad &\leq \sqrt{n} (\sqrt{q})^2 [(E\|A\|^2)^{1/2}]^2 = 2q\sqrt{n} \left( \frac{\sqrt{d} + \sqrt{m}}{\sqrt{m}} \right)^2 \end{aligned}$$

The last inequality comes from Kahane's inequality (see proposition 3.3.1 and proposition 3.4.1 in [11]) and inequality (4.9). Thus, taking  $\varepsilon_{q,m,n,d} = \left(\frac{\sqrt{d}+\sqrt{m}}{\sqrt{m}}\right)^2 \frac{2q}{\sqrt{n}}$ , we have

$$\frac{1}{n} \left\| \left( \sum x_i^2 \right)^{1/2} \right\|_q \leq \varepsilon_{q,m,n,d}. \quad \blacksquare$$

The following lemma is used to prove the first condition in Theorem 4.3.

**Lemma 4.10.** *For  $d \times d$  Wishart random variables  $x_i$ , the following is satisfied*

$$(4.11) \quad \frac{1}{n} \left\| \sum (x_i - Ex_i) \right\|_q \leq \gamma'_q,$$

$$\text{where } \gamma'_q = \begin{cases} C' \ln d \sqrt{\frac{\ln d}{n}} & q \leq \ln d \leq n \\ C' d^{\frac{1}{q}} q \max\{\sqrt{\frac{q}{n}}, \frac{q}{n}\} & q \geq \ln d \end{cases}.$$

*Proof.* By Rosenthal's inequality, we have

$$\begin{aligned}
& \left\| \sum (x_i - Ex_i) \right\|_q \\
& \leq (E \left\| \sum_i (x_i - Ex_i) \right\|_q^q)^{\frac{1}{q}} \\
& \leq_c \sqrt{q} (E |(\sum_i Ex_i^2)^{1/2}|_q^q)^{1/q} + q (\sum |x_i - Ex_i|_q^q)^{1/q} \\
& \leq_c \sqrt{q} (E |(\sum_i Ex_i^2)^{1/2}|_q^q)^{1/q} + q n^{\frac{1}{q}} \cdot \max_i (E |x_i|_q^q)^{\frac{1}{q}} \\
& \leq \sqrt{q} d^{\frac{1}{q}} (E |(\sum_i x_i^2)^{\frac{1}{2}}|_\infty^q)^{1/q} + q n^{\frac{1}{q}} d^{\frac{1}{q}} q \left[ \frac{\sqrt{d} + \sqrt{m}}{\sqrt{m}} \right]^2 \\
& \leq \sqrt{q n} d^{\frac{1}{q}} q \left[ 1 + \sqrt{\frac{d}{m}} \right]^2 + q n^{\frac{1}{q}} d^{\frac{1}{q}} q \left[ 1 + \sqrt{\frac{d}{m}} \right]^2 \\
& \leq d^{\frac{1}{q}} \left[ 1 + \sqrt{\frac{d}{m}} \right]^2 (\sqrt{q n} q + q^2 n^{\frac{1}{q}}).
\end{aligned}$$

The second-to-last inequality uses Kahane's inequality [11] and inequality (4.10). Dividing the inequality by  $n$ , we obtain

$$(4.12) \quad \frac{\left\| \sum (x_i - Ex_i) \right\|_q}{n} \leq d^{\frac{1}{q}} \left[ 1 + \sqrt{\frac{d}{m}} \right]^2 q \left( \sqrt{\frac{q}{n}} + \frac{\sqrt{q}}{n} \sqrt{q n^{\frac{1}{q}}} \right).$$

Let  $2 \leq q_0 \leq q$ . We have two cases to estimate the upper bound:

$$(1) \quad q_0 \leq \ln d \leq n$$

$$(2) \quad \ln d \leq q_0 \leq q$$

We follow the optimization for  $q$  from the proof of Theorem 4.8. Define  $f(q) = \sqrt{q n^{\frac{1}{q}}}$ , and consider  $g(q) = \ln f(q) = \frac{1}{2} \ln q + \frac{1}{q} \ln n$ , then  $g'(q) = \frac{1}{2q} - \frac{\ln n}{q^2} = 0$  at  $q = 2 \ln n$ . Then

$$\frac{\sqrt{q}}{n} f(q) \leq \begin{cases} C \sqrt{\frac{q}{n}} & 2 \leq q < n \\ C \frac{q}{n} & q \geq n \end{cases}$$

Moreover, by (4.12), when  $d \leq m$ , we obtain

$$\begin{aligned}
d^{\frac{1}{q}} \left[ 1 + \sqrt{\frac{d}{m}} \right]^2 q \left( \sqrt{\frac{q}{n}} + \sqrt{\frac{q}{n}} \sqrt{q n^{\frac{1}{q} - \frac{1}{2}}} \right) & \leq 2 C d^{\frac{1}{q}} \left[ 1 + \sqrt{\frac{d}{m}} \right]^2 q \max \left\{ \sqrt{\frac{q}{n}}, \frac{q}{n} \right\} \\
& \leq 8 C d^{\frac{1}{q}} q \max \left\{ \sqrt{\frac{q}{n}}, \frac{q}{n} \right\}.
\end{aligned}$$

Denote  $F(d, n) = 8 C d^{\frac{1}{q}} q \max \left\{ \sqrt{\frac{q}{n}}, \frac{q}{n} \right\}$ . We choose  $q = \ln d$  and we get that  $F(d, n) = C' \ln d \sqrt{\frac{\ln d}{n}}$  if we have  $q_0 \leq \ln d \leq n$ . Otherwise we choose  $q \geq q_0$ , and we get  $F(d, n) = C' d^{\frac{1}{q}} q \max \left\{ \sqrt{\frac{q}{n}}, \frac{q}{n} \right\}$ . Moreover,

$$\hat{\gamma}_q = \begin{cases} C' \ln d \sqrt{\frac{\ln d}{n}} & q \leq \ln d \leq n \\ C' d^{\frac{1}{q}} q \max \left\{ \sqrt{\frac{q}{n}}, \frac{q}{n} \right\} & q \geq \ln d \end{cases}.$$

We apply the estimate for  $q \geq p$  and appeal to Theorem 4.3 and Corollary 4.4.  $\blacksquare$

Now, we can prove the AGM inequality for random matrices which holds up to  $(1 + \varepsilon)$ .

**Theorem 4.11.** *Let  $\{x_i\}$  be a family of self-adjoint family of  $d \times d$  Wishart random matrices. For  $2 \leq p \leq \ln d \leq n$ , we have*

- (i)  $\left\| \sum_i (x_i - E(x_i)) \right\|_p \leq \gamma_p n;$
  - (ii)  $\frac{1}{n} \sum_{i=1}^n E(x_i) = 1;$
  - (iii)  $\left\| \left( \sum_i x_i^2 \right)^{\frac{1}{2}} \right\|_p \leq \gamma_p n$ , where  $\gamma_p = C' \ln d \sqrt{\frac{\ln d}{n}}$ ,  $p_0 \leq \ln d \leq n$ ;
  - (iv) Moreover, for  $\varepsilon \geq 0$  if  $\gamma_p \leq \frac{\varepsilon}{3k}$ ,  $p_k := \frac{p}{k}$  then the following hold.
    - $\|P_k(x_1, \dots, x_n) - EP_k(x_1, \dots, x_n)\|_{p_k} \leq \varepsilon.$
    - The random AGM inequality holds,
- $$\|P_k(x_1, \dots, x_n)\|_{p_k} \leq (1 + 2\varepsilon).$$

*Proof.* Condition (ii) comes from definition of the Wishart random matrices. For condition (i) we directly use Lemma 4.10 for the case when  $p_k \leq \ln d \leq n$ . For condition (iii), we use Lemma 4.9. This implies that all the conditions of Theorem 4.4 are satisfied, since  $p_k \leq \ln d \leq n$ . Thus, we get the random AGM inequality. ■

## 5. APPLICATION ON PISIER'S CONSTRUCTION AND FREELY INDEPENDENT

Let  $(M, \tau)$  be a von Neumann algebra where  $\tau$  is a faithful normal and normalized trace. An example of a finite von Neumann algebra is given by the group von Neumann algebra  $L(G)$  associated to the left regular representation  $\lambda(G)$  of a discrete group  $G$ . It is defined as the strong operator closure of the linear span of  $\lambda(G)$ . Recall that  $L_p(M, \tau)$  where  $1 \leq p < \infty$  is defined as the completion of  $M$  with respect to the norm  $\|x\|_p = (\tau(|x|^p))^{1/p}$  (see Pisier [18] for more details). Note that  $L(G) = L_\infty(L(G))$  and  $L(G) \subset L_p(L(G))$ . We want to prove a version of the AGM inequality with respect to the norm  $\|\cdot\|_p$ . For this version of the AGM inequality, we need the following key lemma.

**Lemma 5.1.** *Let  $M$  be a von Neumann algebra. Let  $\nu$  be a partition. Then there exists a group  $G$  and  $b_i(j) \in L(G)$  such that for  $x_i(j) \in L_p(M)$ , the elements  $X_i(j) = b_i(j) \otimes x_i(j) \in L_p(L(G) \otimes M)$  satisfy*

$$(5.1) \quad [\nu] = E_M \sum_{i_1, i_2, i_3, \dots, i_d} X_{i_1}(1) X_{i_2}(2) \dots X_{i_d}(d).$$

Moreover,

$$(5.2) \quad \left\| \sum_i X_i(j) \right\|_p \leq \begin{cases} C \max \left\{ \left\| \left( \sum x_i(j)^* x_i(j) \right)^{1/2} \right\|_p, \left\| \left( \sum x_i(j) x_i(j)^* \right)^{1/2} \right\|_p \right\} & j \in A_{n.s} \in \sigma_{n.s} \\ \left\| \sum x_i(j) \right\|_p & \{j\} \in \sigma_s, \end{cases}$$

where  $C$  is a universal constant. Note that  $b_i(j) = 1$  if  $\{i\} \in \sigma_s$ .

**Remark 5.2.** The norm inequality (5.2) was proved by Pisier for even integers  $p \geq 2$  in [18]. The general case follows from [8].

Now we can state the AGM inequality for  $L_p(M)$  where  $p \geq d$ .

**Theorem 5.3.** *Let  $M$  be a von Neumann algebra and  $x_i \in L_p(M, \tau)_{s.a}$  satisfy the following condition for some  $\delta \geq 0$ ,*

$$\|(\sum x_i^2)^{1/2}\|_p \leq \delta \|\sum x_i\|_p.$$

*Then we have*

$$(5.3) \quad \|P_d(x_1, \dots, x_n)\|_{\frac{p}{d}} \leq \left(1 + (\delta C)(d! - 1)\right) \frac{n^d(n-d)!}{n!} \left\| \frac{1}{n} \sum_1^n x_i \right\|_p^d.$$

We will only give the sketch of the proof of this theorem since it is similar to the proof of Theorem 3.2 for  $p_d = \frac{p}{d} \geq 1$ .

*Proof.* By using Lemma 5.1, Hölder's inequality and the contraction of conditional expectation we have that

$$\begin{aligned} \|\langle \sigma \rangle\|_{p_d} &\leq \left\| \sum x_i \right\|_p^d + \sum_{v \not\geq \dot{0}} |\mu(\dot{0}, v)| C^{|v_{n.s}|} \|(\sum x_i^2)^{1/2}\|_p^{|v_{n.s}|} \left\| \sum x_i \right\|_p^{|v_s|} \\ &\leq \left\| \sum x_i \right\|_p^d + \sum_{v \not\geq \dot{0}} |\mu(\dot{0}, v)| (\delta C)^{|v_{n.s}|} \left\| \sum x_i \right\|_p^d. \end{aligned}$$

Thus for  $\delta C \leq 1$

$$(5.4) \quad \|P_d(x_1, \dots, x_n)\|_{p_d} \leq (1 + (\delta C)(d! - 1)) \frac{n^d(n-d)!}{n!} \left\| \frac{1}{n} \sum x_i \right\|_p^d. \quad \blacksquare$$

**Remark 5.4.** *If  $\delta \leq 1$ , we get the AGM inequality with a constant  $C(d, n) = C^d d^d$ .*

As a matter of completeness, we want to include the limit case of the Wishart random matrices as an application for the AGM inequality. Let's first give the definition of freely independent von Neumann algebra (for more details see [23]).

**Definition 5.5.** *Let  $\{A_i\}$  be a family of unital von Neumann subalgebras of  $A$ . Then  $\{A_i\}$  is called a freely independent algebra (with respect to a unital linear functional  $\phi$ ) if  $\phi(x_1 \dots x_n) = 0$  whenever  $\phi(x_j) = 0$  for all  $x_j \in A_{i_j}$  and  $i_1 \neq i_2, i_2 \neq i_3, \dots$*

We say that operators  $x_i \in A_i$  are freely independent if their algebra  $\{A_i\}$  are freely independent. In the following theorem we prove the deviation inequality up to  $\varepsilon$  and apply this to the AGM inequality.

**Theorem 5.6.** *If  $\{x_i\}$  are freely independent in von Neumann algebra  $M$  such that*

- (1)  $E_M(x_i) = 1$
- (2)  $\sup_i \|x_i\| \leq C$  and  $2 + (4\sqrt{n})C \leq \frac{\varepsilon n}{3d}$ ,

*then*

- (1)  $\|P_d(x_1, \dots, x_n) - EP_d(x_1, \dots, x_n)\|_\infty \leq \varepsilon$
- (2)  $\|P_d(x_1, \dots, x_n)\|_\infty \leq 1 + \varepsilon$ .



*Proof.* Let  $a_i = x_i - 1$ . By assumption we have  $E_M(a_i) = 0$ . By simple modification of Voiculescu's inequality [7], we get that

$$\begin{aligned}
 \|(\sum a_i^2)^{1/2}\| &= \|\sum e_{i1} \otimes a_i\| \leq \sup \|a_i\| + 2\|(\sum E_M(a_i^2))^{1/2}\| \\
 &\leq 2(1 + C) + 2\sqrt{n}C \\
 (5.5) \quad &\leq 2 + (4\sqrt{n}C) \leq \frac{\varepsilon n}{3d}.
 \end{aligned}$$

Indeed,  $\|a_i\| = \|x_i - 1\| \leq 1 + \|x_i\| \leq 1 + C$  and

$$\begin{aligned}
 E_M(a_i^2) &= E_M(x_i - E_M(x_i))^2 = E_M(x_i^2) - E_M(x_i)^2 \\
 &\leq E_M(x_i^2) = E_M(x_i^{1/2}|x_i|x_i^{1/2}) \\
 &\leq \|x_i\|E_M(x_i) = \|x_i\| \leq C.
 \end{aligned}$$

Again, using Voiculescu's inequality we have,

$$(5.6) \quad \|\sum a_i\| \leq \sup \|a_i\| + 2\|(\sum E_M(a_i^2))^{1/2}\| \leq 2(1 + C) + 2\sqrt{n}C \leq \frac{\varepsilon n}{3d}.$$

Following the proof of Proposition 4.1, we get

$$(5.7) \quad \|\nu\| \leq \|(\sum a_i^2)^{1/2}\|^{d-|v_s|} \|\sum a_i\|^{|v_s|} \leq \left(\frac{\varepsilon n}{3d}\right)^d.$$

Applying the techniques of Proposition 4.1 to the case  $p = \infty$ , we have

$$(5.8) \quad \|P_k(a_1, \dots, a_n) - EP_k(a_1, \dots, a_n)\| \leq 2 \frac{(n-k)!}{n!} k! \left(\frac{\varepsilon n}{3d}\right)^k.$$

Then we have

$$\begin{aligned}
 \|P_d(x_1, \dots, x_n) - EP_d(x_1, \dots, x_n)\| &= \left\| \sum_{k=1}^d \binom{d}{k} (P_k(a_1, \dots, a_n) - EP_k(a_1, \dots, a_n)) \right\| \\
 (5.9) \quad &\leq 2 \sum_{k=1}^d \underbrace{\frac{d!(n-k)!}{(d-k)!n!} n^k}_{f(n) \text{ is a decreasing function}} \left(\frac{\varepsilon}{3d}\right)^k \\
 &\leq 2 \sum_{k=1}^d (d)^k \left(\frac{\varepsilon}{3d}\right)^k \leq \varepsilon.
 \end{aligned}$$

Then we have to apply Theorem 3.7 for  $y_i = Ex_i$  instead of  $x_i$ , where  $\sum_n y_i = 1$ . Note that by free independence, we have  $EP_d(x_1, \dots, x_n) = P_d(Ex_1, \dots, Ex_n)$  using the fact that  $\{x_n\}$  in  $P_d(x_1, \dots, x_n)$  has no repetition.

$$\|P_d(x_1, \dots, x_n)\| \leq \|P_1(Ex_1, \dots, Ex_n)\| + \varepsilon \leq 1 + \varepsilon. \quad \blacksquare$$

**Remark 5.7.** The norm version of the AGM inequality also holds for the family of freely independent  $\{x_i\}$ . Indeed, we have that

$$(5.10) \quad \|P_d(x_1, \dots, x_n)\| \leq (1 + \tilde{\varepsilon}) \left\| \frac{1}{n} \sum x_i \right\|^d.$$

In this case we use again the Voiculescu inequality and deduce that

$$\left\| \frac{1}{n} \sum x_i - \frac{1}{n} \sum Ex_i \right\| \leq \frac{\varepsilon}{3d}.$$

This implies  $\|\frac{1}{n} \sum x_i\| \geq 1 - \frac{\varepsilon}{3d}$ . Hence,

$$\|P_d(x_1, \dots, x_n)\|^{1/d} \leq \frac{(1+\varepsilon)^{1/d}}{(1-\frac{\varepsilon}{3d})} \|\frac{1}{n} \sum x_i\|.$$

Note that  $\frac{(1+\varepsilon)^{1/d}}{(1-\frac{\varepsilon}{3d})} \approx \frac{1+\varepsilon}{1-\varepsilon}$ . This is true since we have

$$(1-x)^n \geq (1-nx)$$

for  $x \in [0, 1]$  and  $n \geq 1$ . Applying this inequality for  $x = \frac{\varepsilon}{3d}$ , we have

$$(1 - \frac{\varepsilon}{3d})^d \geq (1 - \frac{\varepsilon}{3d}d) = (1 - \frac{\varepsilon}{3}).$$

Thus, AGM inequality is true up to the constant  $\frac{1+\varepsilon}{1-\frac{\varepsilon}{3}} \approx 1 + \tilde{\varepsilon}$ .

Another interesting application for freely independent copies  $\{x_i\}$  is given as follows:

**Corollary 5.8.** *Let  $\{x_i\}$  be self-adjoint freely independent copies over an algebra  $B$  such that  $E_B(x_1) = 1_B$  and  $\|x_1\| \leq C$ . Then the AGM inequality holds up to  $(1 + \varepsilon)$ .*

*Proof.* Using the free independence for the  $\{x_i\}$ 's, where  $d \leq p \leq \infty$  we get

- (1)  $E_B(x_i) = 1_B$
- (2)  $\|x_i\|_p = \|x_1\|_p \leq C$
- (3)  $\|(\sum x_i^2)^{1/2}\|_p \leq \tilde{c}\|x_i\|_p n^{1/p} + \sqrt{n}\|(E_M x_i^2)^{1/2}\|_p$

Indeed, for the property (3) we just apply a version of Voiculescu's inequality for free variables [8],

$$\begin{aligned} \|\sum x_i \otimes e_{i1}\|_p &\leq c(\sum \|x_i\|^p)^{1/p} + \|(\sum E_M(x_i^* x_i))^{1/2}\|_p \\ &\leq \tilde{C}n^{1/p} + \sqrt{n}\|(E_M x_1^2)^{1/2}\|_p \end{aligned}$$

Note that

$$\begin{aligned} \|\sum_1^n x_i\|_p &\geq \|\sum_1^n E_M(x_i)\|_p - \|\sum_1^n (x_i - E_M(x_i))\|_p \\ &\geq n\|E_M(x_1)\|_p - \underbrace{\left(\tilde{C}n^{1/p} + \sqrt{n}\|(E_M(x_1^2))^{1/2}\|_p\right)}_A \end{aligned}$$

Now, if  $A \leq \frac{n}{2}\|E_M(x_1)\|_p$ , then we have

$$\begin{aligned} \|(\sum_1^n x_i^2)^{1/2}\|_p &\leq 2 \frac{\tilde{C}n^{1/p} + \sqrt{n}\|(E_M(x_1^2))^{1/2}\|_p}{n\|E_M(x_1)\|_p} \|\sum_1^n x_i\|_p \\ &= n^{-1/2} \underbrace{\left(\frac{2\tilde{C}n^{1/p-1/2} + 2\|(E_M(x_1^2))^{1/2}\|_p}{\|E_M(x_1)\|_p}\right)}_{C_n} \|\sum_1^n x_i\|_p. \end{aligned}$$

Then we get

$$\|(\sum_1^n x_i^2)^{1/2}\|_p \leq \delta_n \|\sum_1^n x_i\|_p,$$

where  $\delta_n = \frac{C}{\sqrt{n}}$ . Then for  $\sqrt{n} \gg d!$  we have  $\delta_n \rightarrow 0$ . This implies that when  $n$  is large enough we get the AGM inequality as follows:

$$(5.11) \quad \|P_d(x_1, \dots, x_n)\|_{\frac{p}{d}} \leq (1 + \varepsilon) \left\| \sum_{i=1}^n x_i \right\|_p^d. \quad \blacksquare$$

## REFERENCES

- [1] George E Andrews. *The theory of partitions*, volume 2. Cambridge university press, 1998.
- [2] Léon Bottou. Online learning and stochastic approximations. *On-line learning in neural networks*, 17(9):25, 1998.
- [3] Xixuan Feng, Arun Kumar, Benjamin Recht, and Christopher Ré. Towards a unified architecture for in-rdbms analytics. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, pages 325–336. ACM, 2012.
- [4] Yehoram Gordon. Some inequalities for Gaussian processes and applications. *Israel Journal of Mathematics*, 50(4):265–289, 1985.
- [5] Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. *Inequalities*. Cambridge university press, 1952.
- [6] Marius Junge. Doob’s inequality for non-commutative martingales. *Journal für die Reine und Angewandte Mathematik*, pages 149–190, 2002.
- [7] Marius Junge. Embedding of the operator space  $oh$  and the logarithmic little Grothendieck inequality. *Inventiones mathematicae*, 161(2):225–286, 2005.
- [8] Marius Junge, Javier Parcet, Quanhua Xu, et al. Rosenthal type inequalities for free chaos. *The Annals of Probability*, 35(4):1374–1437, 2007.
- [9] Marius Junge and Quanhua Xu. Noncommutative Burkholder/Rosenthal inequalities. *Annals of probability*, pages 948–995, 2003.
- [10] Marius Junge, Qiang Zeng, et al. Noncommutative Bennett and Rosenthal inequalities. *The Annals of Probability*, 41(6):4287–4316, 2013.
- [11] Stanislaw Kwapień and Wojbor A Woyczynski. *Random series and stochastic integrals: single and multiple*. Boston, 1992.
- [12] Françoise Lust-Piquard. Inégalités de khintchine dans  $cp(1|p)$ . *CR Acad. Sci. Paris*, 303:289–292, 1986.
- [13] Françoise Lust-Piquard and Gilles Pisier. Non commutative Khintchine and Paley inequalities. *Arkiv för Matematik*, 29(1):241–260, 1991.
- [14] Vitali D Milman and Gideon Schechtman. *Asymptotic theory of finite dimensional normed spaces*. 1986.
- [15] Yu Nesterov. Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM Journal on Optimization*, 22(2):341–362, 2012.
- [16] Vern Paulsen. *Completely bounded maps and operator algebras*, volume 78. Cambridge University Press, 2002.
- [17] Gilles Pisier et al. An inequality for  $p$ -orthogonal sums in non-commutative  $l_{\{p\}}$ . *Illinois Journal of Mathematics*, 44(4):901–923, 2000.
- [18] Gilles Pisier and Quanhua Xu. Non-commutative  $lp$ -spaces. *Handbook of the geometry of Banach spaces*, 2:1459–1517, 2003.
- [19] Narcisse Randrianantoanina. Non-commutative martingale transforms. *Journal of Functional Analysis*, 194(1):181–212, 2002.
- [20] Benjamin Recht and Christopher Re. Parallel stochastic gradient algorithms for large-scale matrix completion. submitted for publication. *Preprint available at <http://pages.cs.wisc.edu/~brecht/publications.html>*, 2011.
- [21] Benjamin Recht and Christopher Ré. Beneath the valley of the noncommutative arithmetic-geometric mean inequality: conjectures, case-studies, and consequences. *arXiv preprint arXiv:1202.4184*, 2012.
- [22] Gian-Carlo Rota. On the foundations of combinatorial theory i. theory of möbius functions. *Probability theory and related fields*, 2(4):340–368, 1964.
- [23] Dan V Voiculescu, Ken J Dykema, and Alexandru Nica. *Free random variables*. Number 1. American Mathematical Soc., 1992.

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

*E-mail address:* `walbar2@illinois.edu`

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

*E-mail address:* `junge@math.uiuc.edu`

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

*E-mail address:* `mzhao16@illinois.edu`